

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES  
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG(19) Weltorganisation für geistiges Eigentum  
Internationales Büro(43) Internationales Veröffentlichungsdatum  
3. Juli 2003 (03.07.2003)

PCT

(10) Internationale Veröffentlichungsnummer  
WO 2003/054189 A3(51) Internationale Patentklassifikation<sup>7</sup>: C12N 15/82,  
A01H 5/00, 5/10

(21) Internationales Aktenzeichen: PCT/EP2002/014302

(22) Internationales Anmeldedatum:  
16. Dezember 2002 (16.12.2002)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:  
101 63 161.8 20. Dezember 2001 (20.12.2001) DE(71) Anmelder (für alle Bestimmungsstaaten mit Ausnahme von  
US): SUNGENE GMBH & CO. KGAA [DE/DE]; Cor-  
rensstrasse 3, 06466 Gatersleben (DE).

(72) Erfinder; und

(75) Erfinder/Anmelder (nur für US): BIESGEN, Christian  
[DE/DE]; Reichenstr. 29, 06484 Quedlinburg (DE).(74) Anwalt: BIEBERBACH, Andreas; BASF Aktiengesellschaft,  
67056 Ludwigshafen (DE).(81) Bestimmungsstaaten (national): AE, AG, AL, AM, AT,  
AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CI, CN, CO, CR,  
CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE,  
GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR,  
KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK,  
MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU,  
SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA,  
UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.(84) Bestimmungsstaaten (regional): ARIPO-Patent (GH,  
GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ,  
TM), europäisches Patent (AT, BE, BG, CH, CY, CZ, DE,  
DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT,  
SE, SI, SK, TR), OAPI-Patent (BF, BJ, CF, CG, CI, CM,  
GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Veröffentlicht:

— mit internationalem Recherchenbericht

(88) Veröffentlichungsdatum des internationalen  
Recherchenberichts: 4. März 2004Zur Erklärung der Zweibuchstaben-Codes und der anderen Ab-  
kürzungen wird auf die Erklärungen ("Guidance Notes on Co-  
des and Abbreviations") am Anfang jeder regulären Ausgabe der  
PCT-Gazette verwiesen.

(54) Title: METHODS FOR THE TRANSFORMATION OF VEGETAL PLASTIDS

(54) Bezeichnung: VERFAHREN ZUR TRANSFORMATION VON PFLANZLICHEN PLASTIDEN

(57) Abstract: The invention relates to novel methods for the production of transgenic plants with genetically modified plastids, in  
addition to transgenic plants produced according to said method.(57) Zusammenfassung: Die vorliegende Erfindung betrifft neue Verfahren zur Herstellung transgener Pflanzen mit gentechnisch  
veränderten Plastiden sowie die mit diesen Verfahren hergestellten transgenen Pflanzen.

WO 2003/054189 A3

## METHODS FOR THE TRANSFORMATION OF VEGETAL PLASTIDS

The present invention relates to novel methods for the generation of transgenic plants with genetically modified plastids and to the transgenic plants generated using these methods.

Biotechnological work carried out on plants aims at generating plants with advantageous novel properties, for example to increase agricultural productivity, to increase the quality in foodstuffs or for producing certain chemicals or pharmaceuticals.

10 Plastids are organelles within plant cells which have their own genome. They play an essential role in photosynthesis and in amino acid and lipid biosynthesis. The plastids' genome consists of a double-stranded, circular DNA with an average size of from 120 to 160 kb and is present - for example in leaf cells - as approximately 1900 to 50 000 copies per cell (Palmer (1985) Ann Rev Genet 19:325-54). A single plastid has a copy number of approximately from 50 to 100. The term plastids comprises chloroplasts, proplastids, etioplasts, chromoplasts, amyloplasts, leukoplasts and elaioplasts (Heifetz P (2000) Biochimie 82:655-666). The various forms can be converted into one another and all arise from the proplastids. This is why all manifested forms of the plastids comprise the same genetic information. Preference is given in the literature as starting material for the transformation of plastids, however, to green cells, which comprise the chloroplasts as the manifested form.

20 It is of great economic interest for plant biotechnologists to develop efficient methods for the transformation of plastids (McFadden G (2001) Plant Physiol. 125:50-53). The stable transformation of plastids of higher plants is one of the great challenges.

In the transformation of plastids, the technique of undirected (illegitimate) DNA insertion, which is frequently employed in insertion into the nuclear DNA, has the disadvantage that it is highly likely that an essential gene on the gene-dense plastidic genome is affected, which would frequently be lethal for the plant. The directed insertion of foreign DNA is therefore advantageous in plastids.

30 Various methods for the directed insertion into the plastidic genome have been described. The first to be described was plastid transformation in green algae (Boynton JE et al. (1988) Science 240: 1534-1538; Blowers AD et al. (1989) Plant Cell 1:123-132),

0817/00028

## 2

followed later by higher plants such as tobacco (Svab Z et al. (1990) Proc Natl Acad Sci USA 87:8526-8530).

EP-A 0 251 654, US 5,932,479, US 5,451,513, US 5,877,402,  
5 WO 01/64024, WO 00/20611, WO 01/42441, WO 99/10513, WO 97/32977,  
WO 00/28014, WO 00/39313 describe methods and DNA constructs for  
the transformation of plastids of higher plants, where the DNA to  
be transformed is introduced into the plastome (plastidic genome)  
via homologous recombination ("double crossover"). In general,  
10 homologous regions of 1000 bp or more on either side of the  
sequence to be inserted are employed. This rapidly gives rise to  
large vectors whose handling is not very convenient. Moreover,  
the transformation efficiency drops. The homologous recombination  
efficiency drops with the increasing length of the foreign DNA to  
15 be integrated. A further disadvantage is the fact that a  
homologous region which can be utilized for the process of DNA  
integration by means of double crossover must be identified for  
each plant species. WO 99/10513 claims the identification of an  
intergenic DNA sequence with supposedly sufficient homology  
20 between the genomes of the chloroplasts of many higher plants,  
which DNA sequence can thus act as a universal target sequence.  
However, it has not been demonstrated that this vector can be  
utilized successfully in species other than tobacco; rather, in  
WO 01/64024, the same inventor adapts the transformation vector  
25 to non-tobacco plant species by using homologous DNA sequences  
isolated from these plant species. Since only few recombination  
events result in all of the above-described methods, selection of  
the recombinant plastidic DNA molecules is required.

30 The plastid DNA of higher plants is present in the form of up to  
several thousand copies per cell. To ensure stable integration of  
foreign DNA, all copies of the plastidic DNA must be modified in  
the same manner. In plastid transformation, this is referred to  
as having reached the homotransplastomic state. This state is  
35 achieved by what is known as a segregation-and-sorting process,  
by exerting a continuous selection pressure on the plants. Owing  
to the continual selection pressure, those plastids in which many  
copies of the plastidic DNA have already been modified are  
enriched during cell and plastid division. The selection pressure  
40 is maintained until the homotransplastomic state is reached (Guda  
C et al. (2000) Plant Cell Reports 19:257-262). The modification  
of all of the copies of the plastidic genome in order to obtain  
homotransplastomic plants which have incorporated the foreign  
gene stably into their plastidic genome over generations without  
45 addition of a selection agent is a great challenge (Bogorad L  
(2000) TIBTECH 18:257-263). In addition to the continuous  
selection pressure, achieving the homotransplastomic state is, if

## 3

appropriate, ensured by repeatedly regenerating tissue which has already been transformed (Svab Z and Maliga P (1993) Proc Natl Acad Sci USA 90:913-917). However, this procedure limits the plant material which is available for plastid transformation.

5 Coupling, if appropriate, the transgene with another gene which is essential for the survival of the plant is therefore proposed. In most cases, tissue culture techniques and selection processes cannot be applied universally to all plant species and constitute a substantial limitation of plastid transformation, in particular

10 with regard to the applicability of the method to species other than tobacco (Kota M et al. (1999) Proc Natl Acad Sci USA 96:1840-1845). A recently published transformation of tomato plastids is based on modifications in the regeneration and selection scheme (Ruf S et al. (2001) Nature Biotech 19:870-875),

15 which, however, are expensive and time-consuming. Another approach aims at reducing the number of plastids per cell and the DNA molecules per plastid so that fewer DNA molecules have to be modified (Bogorad L (2000) TIBTECH 18:257-263). All of the selection and segregation processes are very time-consuming.

20 WO 99/10513 describes a method in which a plastidic ORI (origin of replication) is localized on the plasmid to be transformed in order to increase, in this manner, the number of copies of the vector to be transformed which are available for integration into

25 the plastidic genome (Guda C et al. (2000) Plant Cell Reports 19:257-262).

The necessity of improving the plastid transformation technique is also mentioned in Heifetz and Tuttle (Heifetz P and Tuttle AM

30 (2001) Curr Opin Plant Biol 4:157-161). WO 00/32799 teaches increasing the efficiency of plastid transformation by employing plants with enlarged plastids. This results in a large plastid surface, through which the DNA to be transformed can enter the plastids with greater ease. However, the mechanism of DNA

35 integration relies, again, on conventional homologous recombination, as was the case in the above-described methods.

A variety of other methods for the sequence-specific integration of DNA - in particular into the nuclear DNA - have been

40 described. A method based on self-splicing group II introns has been described. Self-splicing group II introns are capable of inserting in a sequence-specific fashion, for example into intron-free genes. The sequence-specific hydrolysis of the target DNA is catalyzed by an RNA-protein (ribonucleoprotein) complex.

45 Here, the sequence specificity of the endonuclease function is determined in particular by base pairings being formed between the RNA moiety of the ribonucleoprotein complex and the target



DNA. The use of group II introns as vectors for foreign DNA has been discussed. By modifying certain sequences of a group II intron, it was possible to modify the target specificity of the latter. Also, it was possible to insert further sequences into group II introns without destroying functions of the latter (Yang J et al. (1996) *Nature* 381:332-335; Eickbush TH (1999) *Curr Biol* 9:R11-R14; Matsuura M et al. (1997) *Genes Develop* 11:2910-2924; Cousineau B et al. (1998) *Cell* 94: 451-462). The adaptation to certain target sequences and the determination of the associated rules, however, is laborious and has as yet been elucidated in detail only for the Ll.ltrB intron (Mohr G et al. (2000) *Genes Develop* 14:559-573). Moreover, the retrohoming efficiency was reduced significantly by the modification, and not every single one of the modified introns tested inserted into the desired target DNA. The disadvantage of the technique is that some positions in the nucleotide sequence are fixed, which limits the choice of the target region in the DNA to be transformed (Guo H et al. (1997) *EMBO J* 16:6835-6848). Moreover, the efficiency of the retrohoming process with regard to that of the wild-type intron appears to be diminished. The efficiency of intron insertion at different sites on the genes investigated differed with regard to its level. The work aimed at providing an improved method for the directed insertion of DNA into the nuclear DNA of organisms which permit no efficient homologous recombination (Guo et al. (2000) *Science* 289:452-456). The experiments described have been carried out extrachromosomally both in the prokaryote *E.coli* and in human cells. The applicability to the chromosomal DNA of higher organisms or the applicability to plastidic DNA was neither described nor demonstrated. It was merely proposed to attempt the optimization of this system in such a way that insertion into chromosomal DNA of higher eukaryotes can take place. This system is supposed to be an alternative method for higher eukaryotes which lack efficient homologous recombination (Guo et al. (2000) *Science* 289:452-456). This does not apply to plastids of higher plants, where homologous recombination – at least in the case of individual plastidic DNA molecules – can usually be performed without problems.

Plastid transformation was demonstrated not only in tobacco, but also in potato (Sidorov VA et al. (1999) *Plant J* 19:209-216; WO 00/28014), petunia (WO 00/28014), rice (Khan MS and Maliga P (1999) *Nature Biotech* 17:910-915; WO 00/07431; US 6,153,813), *Arabidopsis* (Sikdar SR et al. (1998) *Plant Cell Reports* 18: 20-24; WO 97/32977) and oilseed rape (WO 00/39313). (Review article: Bogorad L (2000) *TIBTECH* 18:257-263). Transplastomic

## 5

tomato plants have also been described recently (Ruf S et al. (2001) Nature Biotech 19:870-875).

The generation of sequence-specific double-strand breaks with the aid of restriction enzymes in eukaryotic genomes, including plants, has been described (Puchta H (1999) Methods Mol Biol 113:447-451).

WO 96/14408 describes the homing restriction endonuclease I-SceI and various possibilities for its use. An application for inserting DNA sequences into plastidic DNA is not described.

Posfai et al. describe a method for the substitution of genes in the prokaryote E.coli (Posfai G et al. (1999) Nucleic Acids Res 27(22):4409-4415). Here, an intramolecular recombination between the endogenous and the mutated gene takes place in the E.coli genome, which combination is induced by cleaving with the restriction enzyme I-SceI. Recombinations in E.coli proceed markedly more efficiently and, presumably, following different mechanisms than is the case in the nucleus of higher eukaryotes (for example described by Kuzminov A (1999) Microbiol Mol Biol Rev. 63(4):751-813).

"Homing" refers to the phenomenon that two or more copies of a DNA sequence exist in one compartment, where at least one of these two sequences is interrupted by a further DNA sequence, and a copy of the interrupting DNA sequence is subsequently also introduced into the noninterrupted DNA sequence. This phenomenon usually takes the form of intron homing. Here, two or more alleles of one gene exist in one compartment, where at least one of these alleles has no intron. A copy of the intron is subsequently also introduced into the intron-free allele.

Introns in plastidic genes of higher plants have been described (Vogel J et al. (1999) Nucl Acids Res 27:3866-3874; Jenkins BD et al. (1997) Plant Cell 9:283-296; Xu MQ et al. (1990) Science 250: 1566-1570). The splicing of a homologous, unmodified intron with the natural exon regions at an ectopic locus in the plastidic genome has likewise been described (Bock R and Maliga P (1995) Nucl Acids Res 23(13):2544-2547). Experiments of introducing, into plastids of higher plants, heterologous introns which are additionally modified in such a way that they comprise additional genetic information and/or splice in a nonnatural sequence environment have not been carried out as yet.

## 6

Experiments carried out by Eddy and Gold into the homing process in *E. coli* have demonstrated that certain recombination systems are required. The type of the recombination system of the host is a key variable (Eddy SR and Gold L (1992) *Proc Natl Acad Sci USA* 89:1544-1547). It was therefore impossible to assume that the naturally occurring homing process of one organism can be applied at will to another organism, in particular when this process probably does not occur naturally in the latter organism.

- 10 Dürrenberger et al. describe the induction of an intrachromosomal recombination in chloroplasts of the single-celled green alga *Chlamydomonas reinhardtii* using the I-CreI homing endonuclease (Dürrenberger F et al. (1996) *Nucleic Acid Res* 24(17):3323-3331). The recombination takes place between the endogenous 23S gene and
- 15 a 23S-cDNA which is inserted into the chromosome of an I-CreI deletion strain and which comprises an I-CreI cleavage site. Double-strand breaks are induced by mating the relevant transgenic organism with an organism which naturally expresses I-CreI. At the point in time of the double-strand break, the
- 20 foreign DNA is already inserted into the chromosomal DNA, and recombination takes place intramolecularly and not between two separate molecules.

It has been shown recently that a mobile intron which naturally

25 occurs in *Chlamydomonas reinhardtii* and which also encodes a homing endonuclease can be transformed efficiently into an intron-free copy (Odom OW et al. (2001) *Mol Cell Biol* 21: 3472-3481). In this work, the increase of the transformation rate was dependent on the presence of the homing endonuclease. In the

30 discussion, it is proposed in general terms and without specific suggestions regarding the implementation, to improve plastid transformation by inducing double-strand breaks. To this end, the recognition regions of rare nucleases were initially to be introduced in a first step, and the subsequent integration event

35 was then to take place at the same locus. More detailed suggestions regarding the manner in which the recognition regions are to be introduced, the type of nucleases and recognition regions which can be used, the way in which the first step and the second step can be designed in actual reality, and the like,

40 are not provided. All that has been shown to date is that the introduction of a homologous intron, into plastids of the alga *Chlamydomonas*, by means of the homing endonuclease naturally associated with the mobility of the intron did work. Moreover, the results were generated in an algal species. The

45 abovementioned experiments by Eddy and Gold with *E. coli*, where no mobile group I introns are known, as is the case with the plastids of higher plants, demonstrate that an applicability to

heterologous systems is not readily feasible. It is therefore by no means obvious for the skilled worker to apply the observations on the alga *Chlamydomonas* to higher plants. In contrast, there are a number of suggestions which make such an applicability rather doubtful:

1. Homing systems cannot be applied readily from one system to another (Eddy SR and Gold L (1992) Proc Natl Acad Sci USA 89:1544-1547). The applicability to higher plants is all the more dubious since no homing endonucleases have been identified in those plastidic genomes of higher plants which have already been sequenced ([http://megasun.bch.umontreal.ca/ogmp/projects/other/cp\\_list.html](http://megasun.bch.umontreal.ca/ogmp/projects/other/cp_list.html)). It can therefore be assumed that the introns found in the plastidic genome of higher plants are not mobile, and that no homing mechanism exists naturally in these genomes.
2. *Chlamydomonas* only has one plastid per cell, while in cells of higher plants up to 100 plastids are present per cell.
3. The efficiency of conventional plastid transformation in *Chlamydomonas* exceeds that in higher plants by several orders of magnitude, which suggests that these two systems cannot be compared directly with one another. As regards the regeneration of transplastomic algae or transplastomic plants, the fact that division of the algal plastids is synchronized with the cell cycle, while this is not the case for the plastids of the higher plants, might also play an important role (Sato N (2001) Trends Plant Science 6:151-155).
4. The mechanisms of DNA integration into plastids of *Chlamydomonas* and of higher plants appear to be fundamentally different. Thus, it has been found that inter-specific plastid transformation (where homologous regions are utilized instead of identical sequences) in *Chlamydomonas* leads to a marked reduction of the transformation efficiency, which was, however, not observed in tobacco. This also applies analogously to the distance of a molecular marker on the homologous DNA from the heterologous sequence on the transformation plasmid: the closer the molecular marker to the edge of the target region for integration by means of double crossover, the less frequently it is transferred when transformed into *Chlamydomonas* plastids. In tobacco, multiple recombination mechanisms were observed, but here even molecular markers which were close to the edge of the homologous regions were transferred efficiently into the

plastidic genome during transformation (Kavanagh TA et al. (1999) Genetics 152: 1111-1122 and references cited therein).

5. In *Chlamydomonas*, the plastids of the two parents fuse during hybridization, even in the case of inter-specific hybridization. In *Chlamydomonas*, plastid fusion is a natural process, and the DNA of the plastids too is mixed and undergoes new recombination. This is why mobile introns in the organelles of these organisms make sense. In contrast, in most of the higher plants, the plastids are inherited uniparentally, so that neither mixing of the plastidic DNA results nor recombinations can occur between the maternal and the paternal plastidic DNA. Even in those plant species in which the plastids are inherited biparentally, no plastid fusion was observed. It can therefore be assumed that natural plastid fusion in higher plants can be ruled out (Hagemann R (1992) plastidic genetics in higher plants; in Cell Organelles, editor: Herrmann RG, Springer Verlag, Vienna, pp.65-96) and that mechanisms like intron homing are either not developed or even suppressed.

- Increasing the homologous recombination efficiency within the nuclear DNA with the aid of rare endonucleases has been described for various organisms (Puchta H et al. (1993) Nucleic Acids Research. 21(22):5034-40; Puchta H et al. (1996) Proc Natl Acad Sci USA 93:5055-5060; Rong YS and Golic KG (2000) Science 28:2013-2018; Jasin M (1996) Trends Genet 12: 224-229). In contrast to plastids, insertion by homologous recombination into the nuclear DNA is problematic and usually takes place owing to random illegitimate integration. This demonstrates that techniques which are established for the nuclear genome cannot necessarily be applied to the plastids. In contrast to the situation regarding the nucleus, integration in plastids of higher plants takes place virtually exclusively, and with high efficiency, via homologous recombination (Bock R and Hagemann R (2000) Progress in Botany 61:76-90; Maliga P et al. (1994) Homologous recombination and integration of foreign DNA in plastids of higher plants. In Homologous recombination and gene silencing in plants. Paszkowski J, ed. (Kluwer Academic publishers), pp.83-93).

- The homologous recombination efficiency for the integration of DNA into the plastome has generally not been thought of as a limiting factor and, in contrast, considered as not being critical. Accordingly, current research into the optimization of plastid transformation does not focus on the optimization of homologous recombination but for example on improved selection

markers, improved selection and regeneration techniques and the like. Nevertheless, no essential breakthrough has been achieved to date.

- 5 As emphasized clearly by the above-described methods and problems in the transformation of plastids, providing novel methods for the generation of homotransplastomic plants is a long-existing, unmet need of plant biotechnology. A further need is the avoidance of antibiotic or herbicide selection markers for  
10 reasons of registration and consumer acceptance. To date, no plastid transformation method has been described which does away with the need for such a selection marker.

It is therefore an object to develop novel methods which ensure  
15 efficient integration of foreign DNA in all copies of the plastidic DNA and which make possible the efficient selection of corresponding homotransplastomic plants. Surprisingly, this object has been achieved by providing the integration/selection method according to the invention.

20

A first subject matter of the invention relates to a method for the integration of a DNA sequence into the plastidic DNA of a multi-celled plant or cell derived therefrom and for the selection of predominantly homotransplastomic cells or plants,  
25 wherein

- a) the plastidic DNA molecules of said multi-celled plant or cell derived therefrom comprise at least one recognition sequence for the directed induction of DNA double-strand  
30 breaks and
- b) at least one enzyme suitable for the induction of DNA double-strand breaks at the recognition sequence for the directed induction of DNA double-strand breaks and at least  
35 one transformation construct comprising an insertion sequence are combined in at least one plastid of said multi-celled plant or cell derived therefrom, and
- c) DNA double-strand breaks are induced at the recognition  
40 sequences for the directed induction of DNA double-strand breaks, and
- d) the insertion sequence inserts into the plastidic DNA, the functionality of the recognition sequence for the directed  
45 induction of DNA double-strand breaks being deactivated so that said recognition sequence is no longer capable of being

## 10

cleaved by the enzyme suitable for the induction of DNA double-strand breaks, and

- e) plants or cells in which the insertion sequence has been  
5 inserted into the plastidic DNA molecules are isolated.

Surprisingly, the system makes possible a substantial increase of the efficiency in the generation of predominantly homotransplastomic plants. In this context, not only the efficacy  
10 of insertion into the plastidic DNA, but also the efficacy of the selection process of predominantly homotransplastomic plants are increased.

Application of the method according to the invention results in a  
15 selection pressure of incorporating the insertion sequence into all of the copies of the plastidic DNA. Ideally, the insertion sequence is spread independently of the selection markers, such as herbicide or antibiotic resistances. This has pronounced advantages with regard to registration and/or consumer  
20 acceptance. However, the use of such selection markers can further increase the efficiency. The method according to the invention for the generation of homoplastomic plants clearly outperforms the prior-art methods since it shows a more rapid, more efficient and therefore more economical route to obtain  
25 homotransplastomic plants. A further advantage of the system is that the size of the constructs employed for the transformation can be kept small since the homologous regions in the plastid transformation vector can be smaller in comparison with the integration by means of double crossover, or can be completely  
30 absent.

The transformation of plastids has a large number of advantages over the transformation of the nucleus. The following are to be mentioned inter alia:

35

- a) While homologous recombination into the nuclear DNA can only be realized with difficulty, DNA in plastids can be integrated readily at a predefined locus by means of double crossover, a form of homologous recombination. Positional  
40 effects or gene silencing, which are encountered in transformations of the nucleus owing to the illegitimate integration at a non-predefined locus, are thus avoided.
- b) Very high expression levels can be achieved, presumably owing  
45 to the high copy number of the plastidic DNA.

## 11

c) In higher plants, plastidic DNA is, as a rule, only subject to maternal inheritance so that the foreign DNA introduced cannot be spread via pollen and cross-pollination can thus be prevented effectively.

5

d) The prokaryotic nature of the plastids makes possible the expression of genes in the context of a polycistronic operon structure. It is therefore not necessary to equate each gene to be expressed with its own promoter and the like. This facilitates the introduction of a large number of genes in one pass, for example for introducing entire biosynthetic pathways into the plastids.

10

"Plastid" refers to the proplastids and to all organelles to which they give rise, such as, for example, chloroplasts, etioplasts, chromoplasts, amyloplasts, leukoplasts, dermaplasts and elaioplasts (Heifetz P (2000) Biochimie 82:655-666).

"Plastome" refers to the genome, i.e. the totality of the genetic information, of a plastid.

20

"Homotransplastomic" refers to a transplastomic and homoplastomic state.

With regard to, for example, a plant, cell, tissue, plastid or a plastidic DNA, "transplastomic" refers to all those forms of the above, realized by recombinant methods, which comprise a plastidic DNA which has been modified by recombinant methods, it being possible for the modification to comprise, for example, substitutions, additions, deletions, inversions or insertions of one or more nucleotide residues.

25

30

"Heteroplastomic" refers to the presence of a mixed population of a variety of plastidic DNAs within a single plastid or within a population of plastids within a plant cell or tissue.

35

"Homoplastomic" refers to a uniform population of plastidic DNA within a single plastid or within a population of plastids within a plant cell or tissue. Homoplastomic cells, tissues or plants are genetically stable since they only comprise one type of plastidic DNA, i.e. they generally remain homoplastomic even when the selection pressure ceases. Progeny obtained by selfing are likewise homoplastomic.

40

For the purposes of the present invention, "predominantly homoplastomic" or "predominantly homotransplastomic" refers to all those plants or cells in which the percentage of the desired

45



## 12

plastidic DNA molecules which have been modified with regard to a trait - for example with the recognition sequence for the directed induction of DNA double-strand breaks or the inserted insertion sequence - amounts to at least 50%, preferably at least 5 70%, very especially preferably at least 90%, most preferably at least 95% of the totality of all plastidic DNA molecules in a plant or a tissue, cell or plastid of same. Predominantly homoplastomic or predominantly homotransplastomic plants can be converted into homoplastomic or homotransplastomic plants by 10 continued maintenance of the selection pressure and, if appropriate, repeated regeneration steps. Owing to the homing process, however, a continuous selection pressure is not necessarily required. In a particular embodiment, a predominantly homoplastomic, or homotransplastomic, plant is therefore truly 15 homoplastomic, or homotransplastomic. A plant which, with regard to a DSB recognition sequence, is predominantly homoplastomic or homotransplastomic, or truly homoplastomic or homotransplastomic, is subsequently referred to as "master plant". The percentage of the desired plastidic DNA molecules which have been modified with 20 regard to a trait can be determined in the manner known to the skilled worker, for example by means of Southern analysis as described by way of example in Example 4. The ratio between the plastid starting DNA molecules and the plastidic DNA molecules which have been modified with regard to a trait can be determined 25 by comparing the intensity of the bands in question.

"Multi-celled plant or cell derived therefrom" refers generally to all those cells, tissues, parts or propagation materials (such as seeds or fruits) of a plant which constitutes, or may 30 constitute, a multi-celled organism in its adult state. Included for the purpose of the invention are all genera and species of higher and lower plants of the plant kingdom. Annual, perennial, monocotyledonous and dicotyledonous plants are preferred. Included are mature plants, seeds, shoots and seedlings, and 35 parts derived therefrom, propagation material (for example tubers, seeds or fruits) and cultures, for example cell or callus cultures. "Mature plants" means plants at any developmental stage beyond the seedling stage. The term seedling means a young immature plant an an early developmental stage.

40 Preferred plants are those from the following plant families: Amaranthaceae, Asteraceae, Brassicaceae, Carophyllaceae, Chenopodiaceae, Compositae, Cruciferae, Cucurbitaceae, Labiatae, Leguminosae, Papilionoideae, Liliaceae, Linaceae, Malvaceae, 45 Rosaceae, Rubiaceae, Saxifragaceae, Scrophulariaceae, Solanaceae, Sterculiaceae, Tetragoniaceae, Theaceae, Umbelliferae.

## 13

Preferred monocotyledonous plants are selected in particular from the monocotyledonous crop plants such as, for example, the Gramineae family such as rice, maize, wheat or other cereal species such as barley, millet and sorghum, rye, triticale or 5 oats, and sugar cane, and all grass species.

Preferred dicotyledonous plants are selected in particular from dicotyledonous crop plants, such as, for example,

- 10 - Asteraceae such as sunflower, tagetes or calendula and others,
  - Compositae, especially the genus Lactuca, very particularly the species sativa (lettuce) and others,
- 15 - Cruciferae, particularly the genus Brassica, very particularly the species napus (oilseed rape), campestris (beet), oleracea cv Tastie (cabbage), oleracea cv Snowball Y (cauliflower) and oleracea cv Emperor (broccoli) and other cabbages; and the genus Arabidopsis, very particularly the species thaliana, and 20 cress or canola and others,
  - Cucurbitaceae such as melon, pumpkin/squash or zucchini and others,
- 25 - Leguminosae, particularly the genus Glycine, very particularly the species max (soybean), soybean, and alfalfa, pea, bean or peanut and others,
  - Rubiaceae, preferably the subclass Lamiidae such as, for 30 example, Coffea arabica or Coffea liberica (coffee bush) and others,
  - Solanaceae, particularly the genus Lycopersicon, very particularly the species esculentum (tomato), the genus 35 Solanum, very particularly the species tuberosum (potato) and melongena (aubergine) and tobacco or paprika and others,
  - Sterculiaceae, preferably the subclass Dilleniidae such as, for example, Theobroma cacao (cacao bush) and others, 40
  - Theaceae, preferably the subclass Dilleniidae such as, for example, Camellia sinensis or Thea sinensis (tea shrub) and others,
- 45 - Umbelliferae, particularly the genus Daucus (very particularly the species carota (carrot)) and Apium (very particularly the species graveolens dulce (celery)) and others; and the genus

## 14

Capsicum, very particularly the genus annuum (pepper) and others,

and linseed, soybean, cotton, hemp, flax, cucumber, spinach,  
5 carrot, sugar beet and the various tree, nut and grapevine species, in particular banana and kiwi fruit.

Also encompassed are ornamental plants, useful or ornamental trees, flowers, cut flowers, shrubs or turf. Plants which may be  
10 mentioned by way of example but not by limitation are angiosperms, bryophytes such as, for example, Hepaticae (liverworts) and Musci (mosses); pteridophytes such as ferns, horsetails and clubmosses; gymnosperms such as conifers, cycads, ginkgo and Gnetatae, the families of the Rosaceae such as rose,  
15 Ericaceae such as rhododendron and azalea, Euphorbiaceae such as poinsettias and croton, Caryophyllaceae such as pinks, Solanaceae such as petunias, Gesneriaceae such as African violet, Balsaminaceae such as touch-me-not, Orchidaceae such as orchids, Iridaceae such as gladioli, iris, freesia and crocus, Compositae  
20 such as marigold, Geraniaceae such as geranium, Liliaceae such as dracena, Moraceae such as ficus, Araceae such as philodendron and many others.

Most preferred are Arabidopsis thaliana, Nicotiana tabacum,  
25 Tagetes and Brassica napus and all those genera and species which are used as foods or feeds, such as the above-described cereal species, or which are suitable for the production of oils, such as oil plants, nut species, soybean, sunflower, pumpkin/squash and peanut.

30  
"Enzyme suitable for inducing DNA double-strand breaks at the recognition sequence for the directed induction of DNA double-strand breaks" (hereinbelow referred to as "DSBI enzyme" for "double strand-break inducing enzyme") generally refers to  
35 all those enzymes which are capable of generating, in a sequence-specific manner, double-strand breaks in double-stranded DNA. The following may be mentioned by way of example, but not by limitation:

40 1. Restriction endonucleases, preferably type II restriction endonucleases, especially preferably homing endonucleases as described in detail hereinbelow.

## 15

2. Artificial nucleases such as described in detail hereinbelow, such as, for example, chimeric nucleases, mutated restriction or homing endonucleases, or RNA protein particles derived from mobile group II introns.

5

Both natural and artificially generated DSB1 enzymes are suitable. Preferred are all those DSB1 enzymes whose recognition sequence is known and which can be obtained either in the form of their proteins (for example by purification) or which can be  
10 expressed using their nucleic acid sequence.

The DSB1 enzyme, whose specific recognition sequence is known, is preferably selected in such a way that it has no further functional recognition regions in the plastidic genome, in  
15 addition to the target recognition sequence. Homing endonucleases are therefore very especially preferred (review: Belfort M and Roberts RJ (1997) Nucleic Acids Res 25:3379-3388; Jasin M (1996) Trends Genet 12:224-228; website:  
<http://rebase.neb.com/rebase/rebase.homing.html>; Roberts RJ and  
20 Macelis D (2001) Nucleic Acids Res 29: 268-269). These meet this requirement owing to their long recognition sequences. Owing to the small size of the plastome, however, it is also feasible that DSB1 enzymes with shorter recognition sequences (for example restriction endonucleases) can be employed successfully.

25

In addition to the above-described preferred embodiment, where only a singular recognition sequence for the DSB1 enzyme is present in the plastidic DNA, cases where further, functionally identical, recognition sequences can be employed advantageously  
30 are also feasible. This is the case in particular when the plastome comprises duplicated genes (for example in the form of inverted repeats). Here, integration into all copies is to take place, so that cleavage in all copies is likewise desirable.

35 The sequences which encode such homing endonucleases can be isolated for example from the chloroplast genome of Chlamydomonas (Turmel M et al. (1993) J Mol Biol 232: 446-467). They are small (18 to 26 kD), but have, in their open reading frame (ORF), a "coding usage" which is directly suitable for expression in the  
40 nucleus or plastids of higher plants (Monnat RJ Jr et al. (1999) Biochem Biophys Res Com 255:88-93).

Further homing endonucleases are mentioned in the abovementioned website; homing endonucleases which may be mentioned are, for  
45 example, F-SceI, F-SceII, F-SuvI, F-TevI, F-TevII, I-AmaI, I-AniI, I-CeuI, I-CeuAIIP, I-ChuI, I-CmoE, I-CpaI, I-CpaII, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIVP,

## 16

I-CsmI, I-CvuI, I-CvuAIP, I-DdiII, I-DirI, I-DmoI, I-HspNIP,  
 I-LlaI, I-MsoI, I-NaaI, I-NanI, I-NclIP, I-NgrIP, I-NitI, I-NjaI,  
 I-Nsp236IP, I-PakI, I-PboIP, I-PcuIP, I-PcuAI, I-PcuVI, I-PgrIP,  
 I-PobIP, I-PorI, I-PorIIP, I-PpbIP, I-PpoI, I-SPBetaIP, I-ScaI,  
 5 I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII,  
 I-SexIP, I-SneIP, I-SpomCP, I-SpomIP, I-SpomIIP, I-SquIP,  
 I-Ssp6803I, I-SthPhiJP, I-SthPhiST3P, I-SthPhiS3bP, I-TdeIP,  
 I-TevI, I-TevII, I-TevIII, I-UarAP, I-UarHGPA1P, I-UarHGPA13P,  
 I-VinIP, I-ZbiIP, PI-MtuI, PI-MtuHIP, PI-MtuHIIP, PI-PfuI,  
 10 PI-PfuII, PI-PkoI, PI-PkoII, PI-PspI, PI-Rma43812IP, PI-SPBetaIP,  
 PI-SceI, PI-TfuI, PI-TfuII, PI-ThyI, PI-TliI, PI-TliII.

Preferred in this context are those homing endonucleases whose  
 gene sequences are already known, such as, for example, F-SceI,  
 15 I-CeuI, I-ChuI, I-DmoI, I-CpaI, I-CpaII, I-CreI, I-CsmI, F-TevI,  
 F-TevII, I-TevI, I-TevII, I-AniI, I-CvuI, I-LlaI, I-NanI, I-MsoI,  
 I-NitI, I-NjaI, I-PakI, I-PorI, I-PpoI, I-ScaI, I-SceI,  
 I-Ssp6803I, PI-PkoI, PI-PkoII, PI-PspI, PI-TfuI, PI-TliI.

20 Homing endonucleases which are especially preferably used are  
 those which are found as such naturally, especially preferably  
 those which are found naturally in organelles. Most preferably,  
 the homing endonucleases originate from organisms which live at  
 similar temperatures to plants. Those which are of particular  
 25 interest in this contest are the homing endonucleases identified  
 in yeast and *Chlamydomonas* species. Naturally, it is also  
 feasible to utilize homing endonucleases which are isolated from  
 extremophilic organisms, as long as they are active in the  
 plastids of the plant to be transformed.

30

The following are very especially preferred:

- I-CeuI (Cote MJ and Turmel M (1995) Curr Genet 27:177-183.;  
 Gauthier A et al. (1991) Curr Genet 19:43-47; Marshall (1991)  
 35 Gene 104:241-245; GenBank Acc. No.: Z17234 nucleotides 5102  
 to 5758),
- I-ChuI (Cote V et al. (1993) Gene 129:69-76; GenBank Acc.  
 No.: L06107, nucleotides 419 to 1075),  
 40
- I-CmoeI (Drouin M et al. (2000) Nucl Acids Res 28:  
 4566-4572),

45

## 17

- I-CpaI from *Chlamydomonas pallidostigmatica* (GenBank Acc. No.: L36830, nucleotides 357 to 815; Turmel M et al. (1995) Nucleic Acids Res 23:2519-2525; Turmel, M et al. (1995) Mol Biol Evol 12:533-545; see also SEQ ID NO: 13 and 14)
- 5
- I-CpaII (Turmel M et al. (1995) Mol Biol Evol 12:533-545; GenBank Acc. No.: L39865, nucleotides 719 to 1423),
- I-CreI (Wang J et al. (1997) Nucleic Acids Res 25: 3767-3776; 10 Dürrenberger, F and Rochaix JD (1991) EMBO J 10:3495-3501; GenBank Acc. No.: X01977, nucleotides 571 to 1062),
- I-CsmI (Ma DP et al. (1992) Plant Mol Biol 18:1001-1004)
- 15 - I-NanI (Elde M et al. (1999) Eur J Biochem. 259:281-288; GenBank Acc. No.: X78280, nucleotides 418 to 1155),
- I-NitI (GenBank Acc. No.: X78277, nucleotides 426 to 1163),
- 20 - I-NjaI (GenBank Acc. No.: X78279, nucleotides 416 to 1153),
- I-PpoI encodes on the extrachromosomal DNA in the nucleus of *Physarum polycephalum* (Muscarella DE and Vogt VM (1989) Cell 56:443-454; Lin J and Vogt VM (1998) Mol Cell Biol 25 18:5809-5817; GenBank Acc. No.: M38131, nucleotides 86 to 577). In addition, the longer sequence encoding I-PpoI, which originates from an alternative start codon, may also be used. This sequence comprises the nucleotides 20 to 577 in the sequence of GenBank Acc. No. M38131. It is preferred to use 30 the shorter sequence; however, it can be substituted at any location by the corresponding, longer one. A sequence which is especially preferred for the purposes of the present invention is an artificial sequence which encodes for the same amino acid as the sequence of the nucleotides 86 to 577 35 of the sequence of GenBank Acc. No.: M38131 (see also SEQ ID NO: 5, 11, 12, 70 or 71),
- I-PspI (GenBank Acc. No.: U00707, nucleotides 1839 to 3449),
- 40 - I-ScaI (Monteilhet C et al. (2000) Nucleic Acids Res 28: 1245-1251; GenBank Acc. No.: X95974, nucleotides 55 to 465)
- I-SceI from the mitochondria of bakers' yeast (WO 96/14408; US 5,962,327 Seq ID NO: 1 therein),
- 45

## 18

- Endo SceI (Kawasaki et al. (1991) J Biol Chem 266:5342-5347, identical with F-SceI; GenBank Acc. No.: M63839, nucleotides 159 to 1589),
- 5 - I-SceII (Sarguiel B et al. (1990) Nucleic Acids Res 18:5659-5665),
- I-SceIII (Sarguiel B et al. (1991) Mol Gen Genet. 255:340-341),
- 10 - I-Ssp6803I (GenBank Acc. No.: D64003, nucleotides 35372 to 35824),
- I-TevI (Chu et al. (1990) Proc Natl Acad Sci USA 87:3574-3578; Bell-Pedersen et al. (1990) Nucleic Acids Res 18:3763-3770; GenBank Acc. No.: AF158101, nucleotides 144431 to 143694),
- 15 - I-TevII (Bell-Pedersen et al. (1990) Nucleic Acids Res 18:3763-3770; GenBank Acc. No.: AF158101, nucleotides 45612 to 44836),
- 20 - I-TevIII (Eddy et al. (1991) Genes Dev. 5:1032-1041),
- 25 Commercially available homing endonucleases such as I-CeuI, I-SceI, I-PpoI, PI-PspI or PI-SceI are very especially preferred. Most preferred are I-SceI and I-PpoI. While the gene encoding I-PpoI can be used in its natural form, the gene encoding I-SceI contains an editing locus. Since, in contrast to yeast
- 30 mitochondria, the editing in question is not performed in the plastids of higher plants, an artificial sequence which encodes the I-SceI protein must be employed for the heterologous expression of this enzyme (US 5,866,361).
- 35 In addition to the above-stated homing endonucleases, there are further intron-encoded enzymes which can be found at homologous locations of the genomes of related organisms. As a rule, these homing endonucleases have similar sequence specificity and are therefore equally suited as DSBi enzyme for the introduction of a
- 40 DSB into the plastome at the DSB recognition sequence. The corresponding sequence is thus also recognized by Sob2593c, Clu2593c, Col2593c, Ciy2593c, Hla2593c, Cag2593c, I-CvuI, I-PakI, Tmu2593c, Msp2593c, I-MsoI, Sdu2593c, Mvi2593m, Nol2593m or Aca2593m, in addition to I-CreI. Corresponding sequence is also
- 45 recognized by I-CecI, I-CmoI, I-CellI, I-CpaIII, I-CmuI, I-CluI, I-SobI or I-AstI, in addition to I-CeuI. A corresponding sequence is also recognized by Cbr1931c, Cfr1931c, Cme1931c, Cge1931c,

## 19

Pcr1931c, Msp1931c, Mso1931c, Ptu1931c, Cvu1931m, Msp1931m, Mso1931m, Noll931m, Aca1931m or Snel931b, in addition to I-CpaI. Moreover, introns exist which are inserted at position 1951 of the 23S rDNA (numbering refers to homologous position in the 23S rDNA of *E. coli*). These introns, again, encode putative homing endonucleases which can be used as DSB1 enzymes for the specific cleavage of the plastidic DNA. They include, for example, Cbr1951c, Msp1951c, Mso1951c, Cvu1951m or Aca1951m (Lucas P et al. (2001) Nucl Acids Res 29:960-969).

10

Most preferred are the homing endonucleases of the protein sequences described by SEQ ID NO: 5, 12 or 14. When preparing corresponding expression cassettes, accordingly, nucleic acid sequences are employed which encode a protein as shown in SEQ ID NO: 5, 12 and 14, respectively; especially preferred in this context is the use of the nucleic acid sequences as shown in SEQ ID NO: 11 and 13 or of an expression cassette as shown in SEQ ID NO: 4.

20 The enzymes can be isolated from their source organisms in the manner with which the skilled worker is familiar and/or the nucleic acid sequence encoding them can be cloned. The sequences of a variety of enzymes have been deposited at GenBank (see above).

25

Examples of artificial DSB1 enzymes by way of example are chimeric nucleases which are composed of an unspecific nuclease domain and a sequence-specific DNA binding domain (for example consisting of zinc fingers) (Smith J et al. (2000) Nucl Acids Res 28(17):3361-3369; Bibikova M et al. (2001) Mol Cell Biol. 21:289-297). Thus, for example, the catalytic domain of the restriction endonuclease FokI has been fused with zinc finger binding domains, whereby the specificity of the endonuclease has been defined (Chandrasegaran S & Smith J (1999) Biol Chem 380:841-848; Kim YG & Chandrasegaran S (1994) Proc Natl Acad Sci USA 91:883-887; Kim YG et al. (1996) Proc Natl Acad Sci USA 93:1156-1160). The catalytic domain of the yeast Ho endonuclease, too, has already been conferred a predefined specificity, using the above-described technique, by fusing it with the zinc finger domain of transcription factors (Nahon E & Raveh D (1998) Nucl Acids Res 26:1233-1239).

As mentioned, zinc finger proteins are particularly suitable as DNA binding domain for the purpose of chimeric nucleases. These DNA-binding zinc finger domains can be adapted to match any desired DNA sequence. Suitable methods for the preparation of such zinc finger domains are described and known to the skilled



## 20

worker (Beerli RR et al. (2000) Proc Natl Acad Sci USA 97(4):1495-1500; Beerli RR et al. (2000) J Biol Chem 275(42):32617-32627; Segal DJ and Barbas CF 3rd. (2000) Curr Opin Chem Biol 4(1):34-39; Kang JS and Kim JS (2000) J Biol Chem 5 275(12):8742-8748; Beerli RR et al. (1998) Proc Natl Acad Sci USA 95(25):14628-14633; Kim JS et al. (1997) Proc Natl Acad Sci USA 94(8):3616-3620; Klug A (1999) J Mol Biol 293(2):215-218; Tsai SY et al. (1998) Adv Drug Deliv Rev 30(1-3):23-31; Mapp AK et al. (2000) Proc Natl Acad Sci USA 97(8):3930-3935; Sharrocks AD 10 et al. (1997) Int J Biochem Cell Biol 29(12):1371-1387; Zhang L et al. (2000) J Biol Chem 275(43):33850-33860). Methods for the preparation and selection of zinc finger DNA binding domains with high sequence specificity have been described (WO 96/06166, WO 98/53059, WO 98/53057). Fusing a DNA binding domain thus 15 obtained with the catalytic domain of an endonuclease (such as, for example, the FokI or Ho endonuclease) allows the preparation of chimeric nucleases with any desired specificity which can be employed advantageously as DSB1 enzymes for the purposes of the present invention.

## 20

Artificial DSB1 enzymes with modified sequence specificity can also be prepared by mutating known restriction endonucleases or homing endonucleases by methods known to the skilled worker. The mutagenesis of maturases with the purpose of obtaining a modified 25 substrate specificity is of particular interest, besides the mutagenesis of homing endonucleases. Frequently, maturases share many features with homing endonucleases and, if appropriate, they can be converted into nucleases by carrying out few mutations. This has been shown, for example, for the maturase in the bakers' 30 yeast *bi2* intron. Only two mutations in the maturase-encoding open reading frame (ORF) sufficed to confer a homing endonuclease activity to this enzyme (Szczepanek & Lazowska (1996) EMBO J 15:3758-3767).

35 Further artificial nucleases can be generated with the aid of mobile group II introns and the proteins encoded by them, or parts of these proteins. Many mobile group II introns, together with the proteins encoded by them, form RNA-protein particles which are capable of recognizing, and cleaving, DNA in a 40 sequence-specific manner. Here, the sequence specificity can be adapted to suit the needs by mutating certain intron regions (see hereinbelow) (WO 97/10362).

## 21

The skilled worker is familiar with various methods for introducing a DSBI enzyme into plastids or expressing it therein. The following may be mentioned by way of example, but not by limitation:

5

## a) Nuclear expression using plastidic transit peptides

10 An expression cassette encoding a DSBI enzyme fusion protein can be constructed in the manner known to the skilled worker, introduced into the nucleus and - optionally - integrated stably into the chromosomal DNA. For transport into the plastids, the DSBI enzyme is preferably expressed in fusion with a plastid localization sequence (PLS). Methods for the direct transportation, into the plastids, of proteins which per se are not localized in the plastids, and a variety of PLS sequences, have been described (Klosgen RB and Weil JH (1991) Mol Gen Genet 225(2):297-304; Van Breusegem F et al. (1998) Plant Mol Biol 38(3):491-496). Preferred are those PLS which, after translocation of the DSBI enzyme into the plastids, are cleaved enzymatically from the DSBI enzyme moiety. Especially preferred is the PLS which is derived from the plastidic Nicotiana tabacum transketolase or from another transit peptide (for example the transit peptide of the small Rubisco subunit or of the ferredoxin NADP oxidoreductase, and also isopentenyl-pyrophosphate isomerase-2) or its functional equivalent. Promoters which are suitable for expression in the nucleus are, in principle, all those which make possible an expression in plants. Examples can be found further below. Preferred are constitutive promoters such as the CaMV 35S promoter or the nitrilase-1 promoter of the Arabidopsis nit1 gene (GenBank Acc. No.: Y07648.2, nucleotides 2456 to 4340; Hillebrand H et al. (1998) Plant Mol Biol 36 (1):89-99; Hillebrand H et al. (1996) Gene 170(2):197-200).

35

Preferred PLS sequences are:

- 40 i) the Arabidopsis isopentenyl isomerase (IPP) transit peptide (GenBank Acc. No.: NC\_003074; nucleotides 604657 - 604486)
- ii) transit peptides derived from the small subunit (SSU) of ribulose-bisphosphate carboxylase (Rubisco ssu) from, for example, pea, maize, sunflower or Arabidopsis.
- 45 - Arabidopsis thaliana: GenBank Acc. No.: for example AY054581, AY054552;

## 22

- 5                   -    pea, GenBank Acc. No.: for example X00806,  
                      nucleotides 1086 to 1256; X04334, X04333 (Hand JM  
                      (1989) EMBO J 8(11):3195-206). Especially preferred  
                      in this context are: expression cassette and transit  
                      peptide (pea, rbcS3A) from vector pJIT117 (Guerineau  
                      F (1988) Nucleic Acids Res 16(23):11380. Especially  
                      preferred is the peptide sequence as shown in SEQ ID  
                      NO: 35. Most preferred for the use in constructing  
10                   suitable expression constructs is the nucleic acid  
                      sequence as shown in SEQ ID NO: 34.
- maize, GenBank Acc. No.: for example S42568, S42508
- sunflower, GenBank Acc. No.: Y00431, nucleotides 301  
15                   to 465.
- iii) transit peptides derived from plant fat biosynthesis  
                      genes, such as the transit peptide of the plastid acyl  
                      carrier protein (ACP) (for example the Arabidopsis  
20                   thaliana beta-ketoacyl-ACP synthetase 2; GenBank Acc.  
                      No.: AF318307), stearyl-ACP desaturase,  $\beta$ -ketoacyl-ACP  
                      synthase or acyl-ACP thioesterase (for example A.thaliana  
                      mRNA for acyl-(acyl carrier protein)thioesterase: GenBank  
                      Acc. No.: Z36911).
- 25                   iv) the GBSSI (starch granule bound synthase I) transit  
                      peptide
- v) the transit peptide of the LHCP II genes.
- 30                   Especially preferred is the PLS of the plastidic tobacco  
                      transketolase (SEQ ID NO: 36). To express corresponding  
                      fusion proteins, different PLS nucleic acid cassettes can be  
                      used in the three reading frames as KpnI/BamHI fragments for  
35                   the purposes of the present invention (the translation start  
                      (ATG codon) is localized in the NcoI cleavage site) (pTP09  
                      SEQ ID NO: 37; pTP10 SEQ ID NO: 38; pTP11 SEQ ID NO: 39).
- 40                   A further example of a PLS to be employed advantageously is  
                      the transit peptide of the plastidic Arabidopsis thaliana  
                      isopentenyl-pyrophosphate isomerase-2 (IPP-2) (SEQ ID NO:  
                      40). The nucleic acid sequences encoding three cassettes  
                      (corresponding to the three reading frames) of the PLS from  
                      the Arabidopsis thaliana isopentenyl-pyrophosphate  
45                   isomerase-2 (IPP-2) can be used very especially preferably

## 23

(EcoRV/SalI cassettes with ATG; IPP-9 SEQ ID NO: 41; IPP-10 SEQ ID NO: 42; IPP-11 SEQ ID NO: 43).

- 5 The nucleic acids according to the invention can be of synthetic origin or have been obtained naturally or comprise a mixture of synthetic and natural nucleic acid components, or else consist of various heterologous gene segments from a variety of organisms.
- 10 The sequence encoding the transit peptide can comprise all or part of the peptide sequence of the original protein. An accurate determination of the amino acid residues which are essential for the transport is not required as long as the functionality of the PLS - which is the transport into the
- 15 plastid - is ensured and the function of the DSBI enzyme is not entirely destroyed. Very especially preferred are the following PLS sequences:
- 20 PLS1: N-MASSSSLTSLQAILSRSVPRHGSASSSQLSPSSLTFSGLKSNPNITTSRRR  
TPSSAAAAVVRSPAIRASAAATETIEKTETAGS-C (SEQ ID NO: 36).  
Corresponds to the PLS of the tobacco plastidic transketolase.
- 25 PLS2: N-MSASSLFNLPLIRLRSLALSSSFSSFRFAHRPLSSISPRKLPNFRAFSGTA  
MTDTKDGSRVDM-C (SEQ ID NO: 40). Corresponds to the PLS of isopentenyl-pyrophosphate isomerase-2 (IPP-2), the last methionine preferably being the start methionine of the DSBI enzyme.
- 30 The homing endonuclease as shown in SEQ ID NO: 69 is a preferred fusion protein of the native I-Ppo I nuclease and the IPP plastid localization sequence. This protein is preferably encoded by the sequence with the SEQ ID NO: 68.
- 35 For the purposes of the present invention, fusion proteins of PLS and DSBI enzyme come under the term "DSBI enzyme". If a DSBI enzyme is expressed in the nucleus, the DSBI enzyme is preferably understood as meaning a fusion protein of PLS and DSBI enzyme.
- 40 The invention furthermore relates to fusion proteins of DSBI enzymes with plastid localization sequence (PLS), sequences and expression cassettes comprising a fusion protein of a plastid localization sequence (PLS) and a DSBI enzyme under the control of a promoter which is functional in the plant
- 45 nucleus. Such suitable promoters are known to the skilled worker and described further below. The expression cassette

## 24

can comprise further elements such as, for example, transcription terminators and/or selection markers.

## b) Expression in plastids

5

An expression in plastids can also take place by the direct introduction of an expression cassette for the DSBI enzyme into plastids, if appropriate integration into the plastidic DNA, and expression of the DSBI enzyme. In a preferred embodiment, this expression cassette is present in the transformation construct which comprises the insertion sequence.

Promoters which can be employed are, firstly, specific plastid or chromoplast promoters as detailed hereinbelow. However, a directed expression in plastids can also be achieved by using for example a viral, bacterial or bacteriophage promoter, introducing the resulting expression cassette into the plastidic DNA, and then inducing expression by the corresponding viral, bacterial or bacteriophage RNA polymerase. The corresponding RNA polymerase, in turn, can be introduced into the plastids in various ways, preferably by nuclear transformation in the form of a fusion protein with a PLS. Suitable methods have been described (WO 95/16783, WO 97/06250, US 5,925,806). It is preferably introduced into plastids by microinjection, especially preferably by means of particle bombardment.

## c) Introduction in the form of RNA

The DSBI enzyme can also be introduced into plastids by introducing the mRNA – for example mRNA which has been generated in vitro – which encodes the DSBI enzyme via, for example, microinjection, particle bombardment (biolistic methods), or polyethylene glycol- or liposome-mediated transfection. This embodiment is advantageous since no DSBI-enzyme-encoding sequences remain in the plastome or genome in this case. Preferably, the RNA encoding the DSBI enzyme is generated by in-vitro transcription in the manner with which the skilled worker is familiar.

40

## d) Introduction in the form of the protein

The DSBI enzyme can be introduced into plastids directly, for example via microinjection, particle bombardment (biolistic methods) or polyethylene glycol transfection or liposome-mediated transfection. This embodiment is advantageous since no DSBI-enzyme-encoding sequences remain

45

## 25

in the plastome or genome. Such a method is described, for example, by Segal DJ et al. (1995) Proc Natl Acad Sci USA 92:806-810.

- 5 The DSBI enzyme can be introduced into plant cells as a fusion protein with the VirE2 or VirF protein of an agrobacterium and a PLS. Such methods have been described for example for Cre recombinase (Vergunst AC et al. (2000) Science 290:979-982). This embodiment is advantageous since  
10 no DSBI enzyme-encoding sequences remain in the genome.

Of course, combinations of the above-described possibilities are also feasible.

- 15 The expression cassette for the DSBI enzyme is preferably present on the insertion sequence or on the transformation construct.

The DSBI enzyme is preferably introduced, or activated, simultaneously with, or after, the introduction of the insertion  
20 sequence into the plastids. Expression and/or activation at the correct site and the correct point in time can be ensured by various approaches:

a) Inducible expression

25

- The expression of a DSBI enzyme can be controlled using an inducible promoter, preferably a chemically inducible promoter. To this end, for example, the expression cassette which encodes the DSBI enzyme can be transformed stably into  
30 the plastidic or nuclear DNA of a master plant. If it is transformed into the nuclear genome, the subcellular localization must be ensured – as described above – by suitable PLS transit peptides. Shortly before or during the transformation with the insertion sequence or the  
35 transformation construct, the expression of the DSBI enzyme will then be switched on by applying a suitable inductor, which depends on the choice of the inducible system. The skilled worker is familiar with a variety of methods or promoters for induced expression. Chemical compounds or else  
40 physical stimuli such as, for example, increased temperature or wounding and the like can act as stimulus. Various examples are described further below.

b) Inducible activity

45

## 26

The DSBI enzyme can already exist in the plastids of the master plant when the activity is induced by suitable techniques at the selected point in time only, for example by addition of chemical compounds. Such methods have been described for sequence-specific recombinases (Angrand PO et al. (1998) Nucl Acids Res 26(13):3263-3269; Logie C and Stewart AF (1995) Proc Natl Acad Sci USA 92(13):5940-5944; Imai T et al. (2001) Proc Natl Acad Sci USA 98(1):224-228). Fusion proteins of the DSBI enzyme and the ligand binding domain of a steroid hormone receptor (for example the human androgen receptor, or mutated variants of the human estrogen receptor as described therein) are employed in these methods. Induction can be effected with ligands such as, for example, estradiol, dexamethasone, 4-hydroxytamoxifen or raloxifen.

Some of the DSBI enzymes are active in the form of the dimer (homo- or heterodimer) (I-CreI forms a homodimer; I-PpoI forms a homodimer, Flick KE et al. (1998) Nature 394: 96-101). In general, enzymes of the LAGLIDADG family tend to form homodimers when only one LAGLIDADG motif is present per monomer (Jurica MS & Stoddard BL (1999) Cell Mol Life Sci 55:1304-1326; I-CeuI may be mentioned by way of example). A dimerization can be designed to be inducible, for example by substituting the natural dimerization domains by the binding domain of a low-molecular-weight ligand. Addition of a dimeric ligand then brings about the dimerization of the fusion protein. Such inducible dimerization methods are described, as is the preparation of the dimeric ligands (Amara JF et al. (1997) Proc Natl Acad Sci USA 94(20): 10618-10623; Muthuswamy SK et al. (1999) Mol Cell Biol 19(10): 6845-6857; Schultz LW and Clardy J (1998) Bioorg Med Chem Lett 8(1):1-6; Keenan T et al. (1998) Bioorg Med Chem. 6(8):1309-1335).

## 35 c) Cotransfection

The expression cassette encoding the DSBI enzyme is preferably introduced into the plastids simultaneously with the insertion sequence. In this context, the expression cassette for the DSBI enzyme and the insertion sequence may be present on one DNA molecule or else on two separate DNA molecules. Preferably, the two sequences are present together on one DNA molecule, so that the expression cassette is present in the transformation construct comprising the insertion sequence.

## 27

In an especially preferred embodiment, the sequence encoding the DSBI enzyme is removed from the genome of the transformed plasmids after homoplastomic plants have been regenerated. The skilled worker is familiar with a variety of methods for  
5 doing so which are detailed further below.

Some of the above-described DSBI enzymes (in particular homing endonucleases) can have recognition sequences in the intermediate host *E. coli*, which is preferably used. Since, moreover, some  
10 expression cassettes for expression in plastids are also functional in *E. coli*, it is preferred to prevent expression of the DSBI enzyme in *E. coli* in various ways with which the skilled worker is familiar in order to avoid any disadvantageous effects on *E. coli* during amplification of the expression cassette. Thus,  
15 for example, several consecutive codons which occur rarely in *E. coli* (for example the codons AGA and AGG, which encode arginine) can be inserted into the reading frame of the DSBI enzyme. This prevents expression in *E. coli*, but – owing to the different codon usage – continues to make possible expression in the  
20 plastids. As an alternative, promoters which are not active in *E. coli*, but are active in the plastids of higher plants can be used (for example promoters of the nuclear encoded plastidic RNA polymerases [NEP promoters; see hereinbelow]). A preferred method is the use of promoters which are recognized neither by the  
25 plastids nor by *E. coli* (for example viral promoters or bacteriophage promoters) and which only become functional by the simultaneous presence of the corresponding viral/bacteriophage RNA polymerase. Such methods are known to the skilled worker and described hereinbelow. Furthermore, it is feasible to destroy the  
30 relevant DSB recognition sequences in *E. coli* or to use a different host which has no DSB recognition sequences for the DSBI enzyme in question. Moreover, it is feasible and advantageous to have the coding region of the DSBI enzyme in promoterless form for amplification in *E. coli*. In this case, the  
35 sequence which encodes the DSBI enzyme is preferably present on a plasmid which is capable of integration into the plastidic genome of the plant to be transformed. Here, the integration site can be chosen in such a way that the gene encoding the DSBI enzyme comes under the control of a promoter which is naturally present in the  
40 plastome or has been inserted artificially into the plastome, thus resulting in expression of the DSBI enzyme in the plastids. A further preferred embodiment ensures that the gene encoding the DSBI enzyme can later be deleted from the plastome (see hereinbelow). In addition, it is possible to create a linkage  
45 between a promoter and a DSBI enzyme by adding such a promoter in vitro upstream of the open reading frame by means of PCR techniques with which the skilled worker is familiar. The PCR



## 28

- product can then be used for introduction into the plant plastids. Moreover, nonfunctional parts of an expression cassette for a DSBI enzyme can be generated and amplified in *E. coli* when these parts undergo recombination with one another after
- 5 introduction into plant plastids (for example by means of homologous recombination in overlapping regions of the nonfunctional moieties of the expression cassette), thus giving rise to a functional expression cassette.
- 10 "Recognition sequence for the directed induction of DNA double-strand breaks" (hereinbelow "DSB recognition sequence" for double-strand break recognition sequence) generally refers to those sequences which permit recognition and cleavage by a DSBI enzyme under the conditions in the plastids of the plant cell or
- 15 plant used in each case. Especially preferred are DSB recognition sequences for homing endonucleases which are encoded naturally in mitochondria or the nucleus of other organisms. Also, it is possible to use DSB recognition sequences of homing endonucleases which are derived from plastids (for example from green algae).
- 20 Preferably, the DSB recognition sequence is singular in the plastidic DNA, i.e. a double-strand break is only generated at the location thus predefined. However, cases where more than one DSB recognition sequence is present in the plastome are also feasible. This is the case in particular when the DSB recognition
- 25 sequence is localized in duplicated genes (for example in inverted repeats). In the latter case, more than one identical DSB recognition sequence exist, but their context is identical, so that, again, directed insertion takes place. Indeed, it is preferred that integration into all copies takes place, which
- 30 also requires cleavage in all copies. DSB recognition sequences which, while occurring more than once in one plastome, are localized in the same plastomic context (for example in repeats or in gene duplications) come under the term "singular DSB recognition sequences" for the purposes of the present invention.
- 35 Preferably, the plant employed, or the cell derived therefrom, is predominantly homoplastomic or homotransplastomic with regard to the DSB recognition sequence, i.e. the predominant number of the plastidic DNA molecules present in the plastid contain this DSB
- 40 recognition sequence. For the purposes of the present invention, such plants are also referred to as master plants.

## 29

In principle, two types of DSB recognition sequences can be used:

a) Natural, endogenous DSB recognition sequences

5 As has been demonstrated within the scope of the present  
invention, the plastomes of higher plants comprise various  
sequences which can act as recognition sequences for DSB  
enzymes (for example homing endonucleases), even though no  
10 such endonucleases have been demonstrated in higher plants to  
date. Such DSB recognition sequences can be identified by  
screening the plastidic DNA sequence using the known DSB  
recognition sequences (for example those described in Table  
2). The plastidic genome of various plants is known  
([http://megasun.bch.umontreal.ca/ogmp/projects/other/cp\\_list.html](http://megasun.bch.umontreal.ca/ogmp/projects/other/cp_list.html)).  
15 The sequences of the plastomes of the following have  
been reported:

- 20 - *Arabidopsis thaliana* (Sato S et al. (1999) DNA Res. 6  
(5):283-290) (GenBank Acc. No.: AP000423; NCBI Acc. No.  
NC\_000932)
- *Epifagus virginiana* (Beechdrops; Wolfe KH et al. (1992) J  
Mol Evol 35(4):304-317; NCBI Acc. No.: NC\_001568; GenBank  
Acc. No.: M81884)
- 25 - *Lotus japonicus* (Kato T et al. (2000) DNA Res 7(6):323-  
330; NCBI Acc. No.: NC\_002694; GenBank Acc. No.: AP002983)
- *Oryza sativa* (rice) (Hiratsuka J et al. (1989) Mol Gen  
30 Genet 217(2-3):185-194; NCBI Acc. No.: NC\_001320; GenBank  
Acc.-No: X15901),
- *Marchantia polymorpha* (Liverwort; Ohyama K et al. (1988)  
J Mol Biol 203(2):281-298; Yamano Y et al. (1984) Nucl  
35 Acids Res 12(11):4621-4624; GenBank Acc. No.: X04465 and  
Y00686; NCBI Acc. No.: NC\_001319)
- *Nicotiana tabacum* (tobacco) (GenBank Acc. No.: Z00044 and  
S54304; NCBI Acc. No.: NC\_001879; Shinozaki K et al.  
40 (1986) EMBO J 5:2043-2049)
- *Oenothera elata* ssp. *hookeri* (Monterey evening primrose;  
GenBank Acc. No.: AJ271079; NCBI Acc. No.: NC\_002693;  
Hupfer H et al. (2000) Mol Gen Genet 263(4):581-585)
- 45 - *Medicago truncatula* (Gen Bank Acc. No.: AC093544)

## 30

- *Pinus thunbergii* (black pine; Tsudzuki J et al. (1994) Curr Genet 26(2):153-158; NCBI Acc. No.: NC\_001631; GenBank Acc. No.: D17510)
- 5     - *Spinacia oleracea* (GenBank Acc. No.: AJ400848 J01442 M12028 M16873 M16878 M27308 M55297 X00795 X00797 X01724 X04131 X04185 X05916 X06871)
- 10    - *Triticum aestivum* (wheat; GenBank Acc. No.: AB042240; NCBI Acc. No.: NC\_002762) and
- *Zea mays* (GenBank Acc. No.: X86563; NCBI Acc. No.: NC\_001666)
- 15    In addition, further plastomes can be sequenced in order to identify DSB recognition sites therein. In general, it suffices to isolate highly-conserved regions from the plastome by PCR methods with which the skilled worker is familiar and to sequence these regions only.
- 20    Furthermore, it is possible to determine natural, endogenous DSB recognition sites experimentally, for example by isolating the plastidic DNA (for example by the method of Mariac P et al. (2000) BioTechniques 28:110-113), amplifying
- 25    the plastidic genome fragments to be taken into consideration by means of PCR or by using synthetic fragments and carrying out a restriction analysis with the DSB1 enzyme in question. This restriction analysis is preferably carried out under conditions as they prevail in the plastid of a higher plant.
- 30    Moreover, the endogenous DSB recognition sequences for natural homing endonucleases which have been identified and described in Table 1 within the scope of the present invention are located in the conserved regions of the
- 35    plastome so that - in particular taking into consideration the given variability, with regard to their respective recognition sequences, of the homing endonucleases mentioned in each case - it can be assumed that these recognition sequences are found virtually universally in all the
- 40    plastomes of higher plants. The positions shown in Table 1 reveal in each case the sequence stated and the reverse-complementary sequence, since all of the recognition regions shown in Table 1 are localized in the inverted repeat of the plastidic genome. Homing endonucleases which
- 45    are especially preferred among those mentioned in Table 1 are I-CpaI, I-CeuI, I-ChuI, I-CpaII and I-CreI.

## 31

The recognition sequences identified thus can be used for the insertion of foreign DNA by generating a double-strand break by introducing the corresponding DSB1 enzyme. If the DSB recognition sequence were to be located in a highly-conserved region within a gene of the organelle genome, the foreign DNA is preferably inserted in the form of a self-splicing intron, which allows the reconstitution of the mRNA of the affected gene (see hereinbelow).

The skilled worker is furthermore familiar with methods in which any endogenous sequence can act as recognition sequence for chimeric, mutated or artificial endonucleases, by subjecting their DNA recognition region to directed modification, for example by modification of a zinc finger domain fused to an endonuclease domain, or by modification of the RNA sequence of a group II intron RNA/protein complex (see hereinabove; WO 96/06166, Bibikova M et al. (2001) Mol Cell Biol 21:289-297).

20

25

30

35

40

45

5

10

15

20

25

30

35

40

45

DSBI enzyme	Published DSB recognition sequence	Sequence in the plastome	Position in the tobacco plastome ACC. 200044	Position in the wheat plastome ACC. AB042240	Position in the rice plastome ACC. X159019	Position in the maize plastome ACC. X86563	Position in the Arabidopsis plastome ACC. AP000423
I-DmoI	ATGCGGCCGGAAC TACCGGCGAGGCAT	GTGCGGGTCGGAACCT ACCGACAGGAAT	C(108281-108310) 134316-134345	118010-118039 C(96855-96884)	117846-117875 C(97243-97272)	121617-121646 C(101091-101120)	131977-132006 C(106643-106672)
I-CpaI	CGATCCTAAGGTAGC GAATTCA	CGGTCCTAAGGTAGCG AAATTCC	108263-108285 C(134341-134363)	96837-96859 C(118035-118057)	97224-97246 C(117871-117893)	101073-101095 C(121642-121664)	106625-106647 C(132002-132024)
I-CeuI	CGTAACATAACGGT CCTAAGGTAGCGAA	CGTAACATAACGGTC CTAAGGTAGCGAA	C(134346-134374) 108252-108280	C(118040-118068) 96826-96854	C(117876-117904) 97214-97242	101062-101090 C(121647-121675)	C(132007-132035) 106614-106642
I-ChuI	GAAGGTTGGCACCT CGATCGGCTCATC	GAAGGTTGGCACCTC GATGCGGCTCTTC	108832-108861 C(133765-133794)	97405-97434 C(117460-117489)	97792-97821 C(117297-117326)	101641-101670 C(121067-121096)	107194-107223 C(131426-131455)
I-CpaII	CCCGGCTAACTCTGT GCCAG	ATCGGCTAACTCTGTG CCAG	C(139398-139417) 103209-103228	C(123374-123393) 91501-91520	C(123351-123370) 91748-91767	C(127108-127127) 95610-95629	C(137169-137188) 101461-101480
I-CreI	CTGGGTTCAAACGT CGTGAGACAGTTGG	CTGGGTTCAAACGTC GTGAGACAGTTCGG	108925-108954 C(133672-133701)	97498-97527 C(117367-117396)	97885-97914 C(117204-117233)	deviating sequence: ctgggttcagaaac gtcgtgagacgtt cgg C(120975-121003) 101734-101762	107287-107316 C(131333-131362)
I-SceI	TACCCGTGTTATCCCT AGCGTAACT	CAGCCTGTTATCCCTA GAGTAACT	C(108804-108781) 133822-133845	C(97377-97354) 117516-117540	C(97741-97764) 117354-117377	C(101590-101613) 121124-121147	C(107143-107166) 131483-131506

## 33

Tab 1: Preferred endogenous cleavage sites in the plastidic genomes of tobacco, wheat, rice, maize and Arabidopsis.

c=complementary. Acc. No: GenBank Accession Number  
(<http://www.ncbi.nlm.nih.gov/>).

5

Singular cleavage sites of restriction endonucleases also exist in the plastidic genome. However, they are usually located in less highly-conserved regions and can therefore not necessarily be exploited universally in all plant species. The following may be mentioned by way of example:

10

- a) With the sequence GGCCTTTATGGCC the enzyme SfiI has a singular recognition site in the plastidic genome of Arabidopsis (GenBank Acc. No.: AP000423) at position 40846-40858.
- b) In the plastidic genome of maize (GenBank Acc. No.: X86563), there is a singular cleavage site for the enzyme AscI at position 42130-42137, with the sequence GGCGCGCC.
- c) In the plastidic genome of rice (GenBank Acc. No.: X159019), there is a singular cleavage site for the enzyme SgfI at position 77309-77316, with the sequence GCGATCGC, and for the enzyme AscI at position 39776-39783 with the sequence GGCGCGCC.
- d) In the plastidic genome of tobacco (Accession Z00044), there is in each case a singular cleavage site for the enzyme SfiI at position 42475-42487, with the sequence GGCCTTTATGGCC, for the enzyme SgrI at position 78522-78529, with the sequence CACCGGCG, and for the enzyme PmeI at position 120895-120902, with the sequence GTTTAAAC.
- e) In the plastidic genome of wheat (Accession AB042240), there is in each case a singular cleavage site for the enzyme PmeI at position 59331-59338, with the sequence GTTTAAAC, a singular cleavage site for the enzymes NarI, KanI, EheI and BbeI at position 41438-41443, with the recognition sequence GGCGCC, and a recognition region for the enzyme SfiI at position 112656-112668, with the sequence GGCCCAGGGGGCC.

35

30

40

45

All these plants with endogenous, natural DSB recognition sequences constitute, in a manner of speaking, naturally occurring master plants. In them, the DSB recognition sequence is naturally present in homoplastomic form. This

eliminates the need for the introduction and selection of artificial DSB recognition sequences.

b) Artificially introduced DSB recognition sequences

5 The skilled worker realizes that the recognition region for a rare enzyme introduced into a master plant need not be of natural origin. In principle, any recognition sequence of any DSBII enzyme can be inserted at any position of the plastidic DNA. The preparation is preferably carried out using a construct for inserting the DSB recognition sequence (hereinbelow DSBR construct). Preferably, the DSBR construct comprises a selection marker to facilitate the selection of transplastomic plants with the successfully inserted DSB recognition sequence, which selection is required for generating suitable master plants. The skilled worker is familiar with a variety of selection markers which make possible selection of plastids (see hereinbelow). *aadA*, *nptII* or *BADH* are preferred, with *aadA* being especially preferred. Selection is carried out for example with the aid of the "segregation and sorting" process, with which the skilled worker is familiar (described by way of example in Example 4). The selection marker is preferably constructed in such a way that a subsequent deletion from the plastome is made possible. Such methods are known to the skilled worker and described hereinbelow.

Thus, it is preferred first to generate a plant which is homotransplastomic with regard to the inserted DSB recognition sequence and which has a DSB recognition sequence in all or the predominant number of the plastids of the plant in question. Such plants can advantageously be employed as master plants.

35 In addition to the selection marker, the DSBR construct may comprise further sequences. These may contain for example further regulatory elements for the expression of the insertion sequences to be introduced subsequently. In a preferred embodiment, the selection marker introduced within the construct for insertion of the DSB recognition sequence is deleted after obtaining the homoplastomic master plant by methods known to the skilled worker (see hereinbelow).

45 In a preferred embodiment, the DSBR construct comprises, for making possible a site-specific insertion, further flanking sequences at at least one, preferably at both sides of the DSB recognition sequence, which flanking sequences have

## 35

sufficient length and homology with corresponding target sequences in the plastome to ensure site-specific insertion by means of homologous recombination.

5 Owing to the large number of the DSBI enzymes with defined recognition sequences which have been described in the prior art, it is possible, and preferred, to generate master plants which have a plurality of different singular DSB recognition sequences incorporated into their plastidic genome.

10

The recognition sequences for the respective DSBI enzymes listed are mentioned hereinbelow in Table 2 by way of example, but not by limitation.

15 Table 2: Recognition sequences and source organism of the DSBI enzymes ("^" shows the cleavage site of the DSBI enzyme within a recognition sequence.)

20	DSBI enzyme	Source organism	Recognition sequence
	I-AniI	Aspergillus nidulans	5'-TTGAGGAGGTT^TCTCTGTAAATAANNNNNNNNNNNNNNNN 3'-AACTCCTCCAAAGAGACATTATNNNNNNNNNNNNNNNN^
	I-CvuI	Chlorella vulgaris	5'-CTGGGTTCAAACGTCGTGA^GACAGTTTGG 3'-GACCCAAGTTTGCAG^CACTCTGTCAAACC
25	I-CsmI	Chlamydomonas smithii	5'-GTACTAGCATGGGGTCAAATGTCTTTCTGG
	I-CmoI	Chlamydomonas moewusii	5'-TCGTAGCAGCT^CACGGTT 3'-AGCATCG^TCGAGTGCCAA
	I-CreI	Chlamydomonas reinhardtii	5'-CTGGGTTCAAACGTCGTGA^GACAGTTTGG 3'-GACCCAAGTTTGCAG^CACTCTGTCAAACC
30	I-ChuI	Chlamydomonas humicola	5'-GAAGGTTTGGCACCTCG^ATGTCGGCTCATC 3'-CTTCCAAACCGTG^GAGCTACAGCCGAGTA
	I-CpaI	Chlamydomonas pallido-stigmatica	5'-CGATCCTAAGGTAGCGAA^ATTCA 3'-GCTAGGATTCCATC^GCTTTAAGT
35	I-CpaII	Chlamydomonas pallido-stigmatica	5'-CCCGGCTAACTC^TGTGCCAG 3'-GGCCGAT^TGAGACACGGTC
	I-CeuI	Chlamydomonas eugametos	5'-CGTAACTATAACGGTCCTAA^GGTAGCGAA 3'-GCATTGATATTGCCAG^GATTCCATCGCTT
40	I-DmoI	Desulfurococcus mobilis	5'-ATGCCTTGCCGGGTAA^GTTCCGGCGCGCAT 3'-TACGGAACGGCC^CATCAAGGCCGCGGTA
	I-SceI	Saccharomyces cerevisiae	5'-AGTTACGCTAGGGATAA^CAGGGTAATATAG 3'-TCAATGCGATCCC^TATTGTCCCATTATATC 5'-TAGGGATAA^CAGGGTAAT 3'-ATCCC^TATTGTCCCATTA ("Core" sequence)
45	I-SceII	Saccharomyces cerevisiae	5'-TTTTGATTCTTTGGTCACCC^TGAAGTATA 3'-AAACTAAGAAACCAG^TGGGACTTCATAT
	I-SceIII	Saccharomyces cerevisiae	5'-ATTGGAGGTTTGGTAAC^TATTTATTACC 3'-TAACCTCCAAAACC^ATTGATAAATAATGG



	DSBI enzyme	Source organism	Recognition sequence
5	I-SceIV	<i>Saccharomyces cerevisiae</i>	5'-TCTTTTCTCTTGATTA^GCCCTAATCTACG 3'-AGAAAAGAGAAC^TAATCGGGATTAGATGC
	I-SceV	<i>Saccharomyces cerevisiae</i>	5'-AATAATTTTCT^TCTTAGTAATGCC 3'-TTATTAAAAGAAGAATCATTAC^CGG
	I-SceVI	<i>Saccharomyces cerevisiae</i>	5'-GTTATTTAATG^TTTTAGTAGTTGG 3'-CAATAAATTACAAAATCATCA^ACC
10	I-SceVII	<i>Saccharomyces cerevisiae</i>	5'-TGTCACATTGAGGTGCACTAGTTATTAC
	PI-SceI	<i>Saccharomyces cerevisiae</i>	5'-ATCTATGTCGGGTGC^GGAGAAAGAGGTAAT 3'-TAGATACAGCC^CACGCCTCTTTCTCCATTA
	F-SceI	<i>Saccharomyces cerevisiae</i>	5'-GATGCTGTAGGC^ATAGGCTTGGTT 3'-CTACGACA^TCCGTATCCGAACCAA
15	F-SceII	<i>Saccharomyces cerevisiae</i>	5'-CTTTCGCAACA^GTAAAATT 3'-GAAAGGCG^TTGTCATTTTAA
	I-LlaI	<i>Lactococcus lactis</i>	5'-CACATCCATAAC^CATATCATTTTT 3'-GTGTAGGTATTGGTATAGTAA^AAA
	I-MsoI	<i>Monomastix species</i>	5'-CTGGGTTCAAACGTCGTGA^GACAGTTTGG 3'-GACCCAAGTTTTCAG^CACTCTGTCAAACC
20	I-NanI	<i>Naegleria andersoni</i>	5'-AAGTCTGGTGCCA^GCACCCGC 3'-TTCAGACC^ACGGTCGTGGGCG
	I-NitI	<i>Naegleria italica</i>	5'-AAGTCTGGTGCCA^GCACCCGC 3'-TTCAGACC^ACGGTCGTGGGCG
	I-NjaI	<i>Naegleria jamiesoni</i>	5'-AAGTCTGGTGCCA^GCACCCGC 3'-TTCAGACC^ACGGTCGTGGGCG
25	I-PakI	<i>Pseudomonas aeruginosa</i>	5'-CTGGGTTCAAACGTCGTGA^GACAGTTTGG 3'-GACCCAAGTTTTCAG^CACTCTGTCAAACC
	I-PorI	<i>Pyrobaculum organotrophum</i>	5'-GCGAGCCCGTAAGGGT^GTGTACGGG 3'-CGCTCGGGCATT^CCCACACATGCC
30	I-PpoI	<i>Physarum polycephalum</i>	5'-TAACATGACTCTCTTAA^GGTAGCCAAAT 3'-ATTGATACTGAGAG^AATTCATCGGTTTA "Core sequence": CTCTCTTAA^GGTAGC GAGAG^AATTCATCG
	I-ScaI	<i>Saccharomyces capensis</i>	5'-TGTCACATTGAGGTGCACT^AGTTATTAC 3'-ACAGTGTAATCCAC^GTGATCAATAATG
	I-Ssp6803I	<i>Synechocystis species</i>	5'-GTCGGGCT^CATAACCCGAA 3'-CAGCCCGAGTA^TTGGGCTT
40	PI-PfuI	<i>Pyrococcus furiosus</i> Vc1	5'-GAAGATGGGAGGAGGG^ACCGGACTCAACTT 3'-CTTCTACCCCTC^TCCCTGGCCTGAGTTGAA
	PI-PfuII	<i>Pyrococcus furiosus</i> Vc1	5'-ACGAATCCATGTGGAGA^AGAGCCTCTATA 3'-TGCTTAGGTACAC^CTCTTCTCGGAGATAT
	PI-PkoI	<i>Pyrococcus kodakaraensis</i> KOD1	5'-GATTTTAGAT^CCCTGTACC 3'-CTAAAA^TCTAGGGACATGG
45	PI-PkoII	<i>Pyrococcus kodakaraensis</i> KOD1	5'-CAGTACTACG^GTTAC 3'-GTCATG^ATGCCAATG
	PI-PspI	<i>Pyrococcus sp.</i>	5'-AAAATCCTGGCAAACAGCTATTAT^GGGTAT 3'-TTTtaggaccgTTTGTGAT^AATACCCATA

37

	DSBI enzyme	Source organism	Recognition sequence
5	PI-TfuI	<i>Thermococcus fumicolans ST557</i>	5'-TAGATTTAGGT^CGCTATATCCTTCC 3'-ATCTAAAA^TCCAGCGATATAGGAAGG
	PI-TfuII	<i>Thermococcus fumicolans ST557</i>	5'-TAYGCNGAYACN^GACGGYTTYT 3'-ATRCGNCT^RTGNCTGCCRAARA
10	PI-ThyI	<i>Thermococcus hydro- thermalis</i>	5'-TAYGCNGAYACN^GACGGYTTYT 3'-ATRCGNCT^RTGNCTGCCRAARA
	PI-TliI	<i>Thermococcus litoralis</i>	5'-TAYGCNGAYACNGACGG^YTTYT 3'-ATRCGNCTRTGNC^TGCCRAARA
	PI-TliII	<i>Thermococcus litoralis</i>	5'-AAATTGCTTGCAAACAGCTATTACGGCTAT
15	I-TevI	<i>Bacteriophage T4</i>	5'-AGTGGTATCAAC^GCTCAGTAGATG 3'-TCACCATAGT^TGCGAGTCATCTAC
	I-TevII	<i>Bacteriophage T4</i>	5'-GCTTATGAGTATGAAGTGAACACGT^TATTC 3'-CGAATACTCATACTTCACTTGTG^CAATAAG
	F-TevI	<i>Bacteriophage T4</i>	5'-GAAACACAAGA^AATGTTTAGTAAANNNNNNNNNNNNN 3'-CTTTGTGTTCTTTACAAATCATTNNNNNNNNNNNNNN^
20	F-TevII	<i>Bacteriophage T4</i>	5'-TTTAATCCTCGCTTC^AGATATGGCAACTG 3'-AAATTAGGAGCGA^AGTCTATACCGTTGAC

Also comprised are deviations (degenerations) of the recognition sequence which nevertheless continue to make possible recognition and cleavage by the DSBI enzyme in question. Such deviations – also in connection with different framework conditions such as, for example, calcium or magnesium concentrations – have been described (Argast GM et al. (1998) J Mol Biol 280: 345-353).

Furthermore comprised are core sequences of these recognition sequences. It is known that the inner portions of the recognition sequences also suffice for an induced double-strand break and that the outer portions are not necessarily relevant, but may have an effect on the cleavage efficiency. Thus, for example, an 18 bp core sequence can be defined for I-SceI. The term "DSB recognition sequence" thus also comprises all essentially identical recognition sequences. Essentially identical recognition sequences refers to those recognition sequences which, while deviating from the recognition sequence identified as being optimal for the enzyme in question, still permit cleavage by the same.

Various localization sites (in the case of already existing endogenous DSB recognition sequences) or integration sites (in the case of artificially generated DSB recognition sequences) are

possible for the DSB recognition sequence. Examples which may be mentioned are:

- a) Localization (integration) in a transcriptionally silent  
5 region

Localization (integration) of the DSB recognition sequence in  
a transcriptionally silent region of the plastidic genome  
(intergenic region) is the preferred embodiment. In this  
10 manner, an adverse effect on the plastids' functions can be  
largely ruled out. In this context, it must be noted that, if  
appropriate, suitable regulatory elements such as promoters  
and the like must also be introduced for expression to take  
place.

- 15 b) Localization (integration) in a transcriptionally active but  
noncoding (intercistronic) region

The advantage of this localization (integration) is that the  
20 insertion sequence to be introduced is thereby ultimately  
encoded in a plastidic operon and promoter(s)/terminator(s)  
need not be introduced separately, but those present  
endogenously at this locus can, but do not have to, be  
utilized. In such a case, only ribosome binding sites should  
25 be present at a suitable distance upstream of the coding  
region of the foreign genes to be introduced.

However, it is also feasible that an intergenic region is not  
entirely transcriptionally silent, for example because  
30 transcriptional termination from an adjacent gene or operon  
is only inefficient.

- c) Localization (integration) in a transcriptionally active  
35 coding region.

The localization (integration) described under a) and b) of  
the DSB recognition sequence at a noncoding locus has the  
advantage that the insertion of the foreign DNA is highly  
likely not to affect the function of the plastidic genome.  
40 However, noncoding regions are less well conserved than  
coding regions. In order to have available as universal a  
method as possible which works in many plant species, the DSB  
recognition sequence (and therefore the insertion sequence)  
is, in an especially preferred embodiment, localized in the  
45 coding sequence of an existing gene. Destruction of the gene  
function by introducing the DSB recognition sequence (in the  
case of an artificially generated DSB recognition sequence),

## 39

or the introduction of the insertion sequence, is prevented inventively, in a preferred variant of this embodiment, by introducing the DSB recognition sequence, or the insertion sequence, within an intron. In this manner, the complete coding mRNA is regenerated at the site of integration by splicing the pre-RNA of the gene.

DSB recognition sequences which do not occur naturally in the plastidic DNA can be introduced into the plastidic DNA in various ways. Examples which may be mentioned are:

a) Integration by means of double crossover

Integration into the plastidic genome is preferably carried out with the aid of the above-described methods with which the skilled worker is generally familiar (double crossover).

b) Integration using natural, endogenous DSB recognition sequences

20

c) Integration using recombinases and corresponding recognition sequences.

Even though the procedure for inserting an artificial DSB recognition sequence into the plastidic DNA is relatively complicated and, in case a), corresponds to the plastid transformation method currently described in the prior art, this complicated procedure only has to be carried out once. The resulting homotransplastomic master plant can then be employed for any number of different subsequent transformations using the method according to the invention, which makes possible a substantial increase in the transformation efficiency: instead of having to carry out the conventional selection process for a homotransplastomic plant every single time, it only has to be carried out once in the present context.

"Deactivation of the functionality" of a DSB recognition sequence means that, owing to insertion of the insertion sequence at or near the position of the double-strand break, the DSB recognition sequence is destroyed, i.e. the corresponding DSBI enzyme no longer recognizes the region and, accordingly, no longer induces a double-strand break at this position.

Construction of the transformation construct with the insertion sequence

## 40

Using one of the above-described master plants or cells derived from them which contain a natural and/or an artificially generated DSB recognition sequence in the plastome, the insertion sequence is inserted into said DSB recognition sequence within a 5 transformation process. This is effected with the simultaneous presence of a DSB1 enzyme, which recognizes one of the DSB recognition sequences in the plastome.

In its simplest form, the transformation construct consists only 10 of the insertion sequence itself, for example of an expression cassette which is to ensure the expression of a certain gene in the plastids. The sequence-specific induction of double-strand breaks suffices to ensure that this insertion sequence is placed at this position and thus to bring about the deactivation of the 15 DSB recognition sequence.

In a preferred embodiment, the insertion sequence comprises at least one nucleic acid sequence to be expressed. To ensure expression (transcription and/or translation), they are to be 20 provided with regulatory elements, depending on the embodiment and the insertion site. If insertion takes place at a transcriptionally active locus, no promoter sequences are required, as described above. The sequences to be expressed are advantageously provided in any case with ribosome binding sites 25 at a suitable distance upstream of the open reading frame, or are already equipped naturally with such sites. These regulatory sequences or parts thereof can, however, also be present naturally in the plastome or introduced into the plastidic DNA together with the DSB recognition sequence as early as in the 30 first step, i.e. in the generation of a nonnatural master plant.

An increase of the insertion efficiency and insertion accuracy can be brought about by flanking the insertion sequence present in the transformation construct and the DSB recognition sequence 35 by homologous sequence regions which, owing to the induced double-strand break, ensure homologous recombination. In a preferred embodiment, the insertion sequence comprises flanking homology sequences A' and B', the sequence to be introduced into the plastidic DNA being located between A' and B'. The DSB 40 recognition sequence is flanked by homology sequences A and B, respectively, the DSB recognition sequence being located between A and B. A and B can be of natural origin or have been introduced in context with the insertion of nonnatural DSB recognition sequences. A and A' and B and B', respectively, are sufficiently 45 long and sufficiently homologous to one another to ensure a

## 41

homologous recombination between A and A', and B and B', respectively.

In a further embodiment, the DSB recognition sequence is flanked merely by a homologous sequence A which has sufficient homology to a sequence A' which flanks the insertion sequence unilaterally.

With regard to the homology sequences, "sufficient length" preferably means sequences with a length of at least 20 base pairs, preferably at least 50 base pairs, especially preferably at least 100 base pairs, very especially preferably at least 250 base pairs, most preferably at least 500 base pairs.

With regard to the homology sequences A and A', and B and B', respectively, "sufficient homology" preferably means sequences which have at least 70%, preferably 80%, by preference at least 90%, especially preferably at least 95%, very especially preferably at least 99%, most preferably 100% homology within these homology sequences over a length of at least 20 base pairs, preferably at least 50 base pairs, especially preferably at least 100 base pairs, very especially preferably at least 250 base pairs, most preferably at least 500 base pairs.

Homology between two nucleic acids is understood as meaning the identity of the nucleic acid sequence over in each case the entire sequence length which is calculated by alignment with the aid of the program algorithm GAP (Wisconsin Package Version 10.0, University of Wisconsin, Genetics Computer Group (GCG), Madison, USA), setting the following parameters:

Gap Weight: 12

Length Weight: 4

Average Match: 2,912

Average Mismatch: -2,003

Since homologous recombination is promoted by the induced double-strand break, the requirements regarding length and homology of the sequences are markedly less than is the case for example in the case of conventional homologous recombination. In this context, the homologous regions can also be markedly smaller than 250 bp. The advantage of using homology sequences is that, when A' and B' are different, or when only one homology sequence A' is being used, a directed insertion of the insertion sequence into the plastidic DNA can take place.

## 42

The transformation construct or the insertion sequence preferably comprises a selection marker which makes possible the selection of transplastomic plastids (see hereinbelow), especially preferably aadA, BADH or a binding-type marker. The selection marker is preferably constructed in such a way that subsequent deletion from the plastome is made possible. Such methods are known to the skilled worker and described hereinbelow.

The insertion sequence or the transformation construct preferably has the structure and sequence of an intron. As a rule, the naturally occurring introns are modified in such a manner for this purpose that they meet the requirements of the method according to the invention. Such artificial introns are especially preferred when they are to be inserted into a transcriptionally active or even coding region, for example, using a natural, endogenous DSB recognition sequence. Preferably, insertion takes place in such a way that the inserted sequence is removed completely by splicing the pre-mRNA. The RNA which has been spliced out (that is to say the artificial intron) now constitutes the mRNA, for example for the translation of proteins encoded on it. This method has further advantages:

- The introns utilized show pronounced secondary folding so that a relatively stable RNA results. The genes of interest which are encoded in the intron can therefore be expressed at a particularly high level, as has been demonstrated, for example, in *E. coli* (Chan KYW et al. (1988) Gene 73:295-304).
- When the intron is integrated into a gene, the transcription of the intron is subject to the regulatory control of the gene into which the intron has been integrated. This is why all regulatory elements upstream or downstream of the gene(s) of interest can be dispensed with in the intron. The constructs can thus be kept correspondingly small, and it is certain that transcription does indeed work, including in the species under investigation. The utilization of heterologous regulatory elements involves the residual risk that these elements are not functional in the investigated plastids of the plant species in question. The utilization of homologous sequences can, owing to the sequence duplication, lead to spontaneous recombination events with the endogenous sequences and thus to instability of the organelle genome. Owing to the possibility of largely being able to dispense with the introduction of regulatory elements - for example by encoding the gene of interest in an intron which is inserted into a transcriptionally active plastome region - many other disadvantages of conventional plastid transformation can be

## 43

avoided with the method according to the invention in this embodiment, in addition to increasing the insertion and distribution ability of the transformation constructs.

5 Moreover, all introns can be used when the relevant factors which mediate splicing are simultaneously expressed in the plastids or imported into them. Preferably, the splicing factors are encoded in the intron itself. Group II introns, which themselves encode at least one of the splice factors, are especially preferred in  
10 this embodiment. They include the *Lactococcus* Ll.ltrB intron. Likewise preferred introns are those which naturally occur in the plastids of higher plants, especially group II introns, very especially preferably introns which encode a protein, most preferably introns of the trnK genes of the plastidic genome. In  
15 the latter case, the introns from the trnK genes of the plastids from the species *Arabidopsis*, maize and tobacco are especially preferred.

Preferred introns are those which have a self-splicing activity  
20 which does not depend on further protein factors, or introns which utilize general factors for splicing which are universally present, and therefore also in plastids, and also introns which themselves encode factors required for splicing. These introns include, for example,

25

- a) the group I intron from *Tetrahymena* (GenBank Acc. No.: X54512; Kruger K et al. (1982) *Cell* 31:147-157; Roman J and Woodson SA (1998) *Proc Natl Acad Sci USA* 95:2134-2139)
- 30 b) the group II rII intron from *Scenedesmus obliquus* (GenBank Acc. No.: X17375.2 nucleotides 28831 to 29438; Holländer V and Kück U (1999) *Nucl Acids Res* 27: 2339-2344; Herdenberger F et al. (1994) *Nucl Acids Res* 22: 2869-2875; Kück U et al. (1990) *Nucl Acids Res* 18:2691-2697).
- 35 c) the Ll.LtrB intron (GenBank Acc. No.: U50902 nucleotides 2854 to 5345)
- d) the *Arabidopsis* trnK intron (GenBank Acc. No.: AP000423, complementary nucleotides 1752 to 4310)  
40
- e) the maize trnK intron (GenBank Acc. No.: X86563, complementary nucleotides 1421 to 3909)
- 45 f) the tobacco trnK intron (GenBank Acc. No.: Z00044, complementary nucleotides 1752 to 4310).



## 44

Not only heterologous introns, but also introns which naturally occur in the plastids of the plant in question can be utilized. Heterologous introns – for example heterologous trnK introns – are preferred to avoid instabilities brought about by sequence duplication. In a preferred embodiment, introns which occur naturally in the plastids of the plant in question are modified in such a way that they have a sequence homology of less than 95%, preferably 80%, especially preferably 70% with the sequence of the starting intron, while still being able to retain their function.

In a further preferred embodiment, a factor which brings about splicing of the intron in question is available in trans, i.e. it is not encoded in the intron itself. If this factor is not naturally present in the plastid in question, but first has to be introduced into it, such a procedure can be effected in various ways with which the skilled worker is familiar. Examples which may be mentioned are the introduction of a suitable coding sequence, which is capable of expression, into the plastome or the introduction into the nuclear DNA; in the latter case, the factor is preferably fused with a PLS.

Especially preferred introns are those which naturally encode a DSB enzyme (in particular a homing endonuclease). Especially preferred is the intron Cp.LSU2 from *Chlamydomonas pallidostigmatica*, which encodes the enzyme I-CpaI (Turmel M et al. (1995) Mol Biol Evol 12:533-545). Also preferred are the group II introns from yeast mitochondria.

In a preferred embodiment, the intron sequence is adapted to suit the insert site so that they can splice at this locus. In the case of group I introns, this adaptation can relate to the internal guide sequence (IGS) and in the case of the group II introns the exon binding sequence (EBS) I and/or II.

In the case of the maize trnK intron, it must be noted that the protein encoded by the trnK intron, which also comprises the maturase function, is probably not functional in its naturally encoded form without editing. It has been demonstrated that editing (His420Tyr) of the corresponding mRNA takes place in barley plastids (Vogel J et al. (1997) J Mol Biol 270:179-187). Tyrosine at position 420 of the matK protein is highly conserved. In the monocots rice and maize, too, a codon encoding His has been found at the corresponding position in the coding DNA. It can therefore be assumed that the matK transcript is also edited in those plants, as is the case in barley. Since, however, other plant species may, if appropriate, not be able to provide such

## 45

RNA editing, a preferred embodiment provides that the *matK* gene in the maize *trnK* intron is already modified at DNA level by a suitable His/Tyr substitution, so that RNA editing is no longer required. For example, the sequence CATTATCATAGTGGAT of the maize 5 *trnK* intron can be mutated into CATTATTATAGTGGAT.

In the case of group I introns, the splicing site is determined by the pairing of IGS with the exon of the corresponding transcript, which exon is located 5' and/or 3' relative to the 10 intron (Lambowitz AM & Belfort M (1993) *Annu Rev Biochem* 62:587-622). Using techniques which are known to the skilled worker, such as PCR or the synthetic generation of nucleotide sequences, the IGS can be matched to any group I introns in such a way that splicing takes place at the predefined insertion site 15 within the DSB recognition region. The modified IGS is designed in such a way that it can undergo - at least partial - base pairing with the sequences of the transcript 5' and 3' of the insertion site. The *C. pallidostigmatica* *CpLSU2* intron, which encodes the homing endonuclease I-CpaI, is preferably utilized. 20 If this intron is utilized in connection with the expression of the DSB1 enzyme I-CpaI, whereby insertion of the DNA to be transformed into the 23S rDNA of the plastidic genome of higher plants results, no adaptation of the intron is necessary. Insertion takes place at a locus in the plastidic genome of 25 higher plants which is homologous to the locus at which the intron is naturally present in *C. pallidostigmatica*. This intron is therefore already designed in such a way that pairings with the 5' and 3' exon can be undergone and that correct splicing in this nucleotide environment takes place. Furthermore preferred is 30 the group I intron from *Tetrahymena thermophila*, where, as a 413 bp intervening sequence (IVS), it interrupts the 26S rRNA coding region (Accession V01416 J01235 nucleotides 53 to 465). The IGS with the sequence 5'-ggaggg-3' which can be found naturally (Waring RB et al. 1985 *Cell* 40: 371-380; Been, MD & Cech, 35 *TR* 1986 *Cell* 47: 207-216) can be adapted to the new insertion site by techniques with which the skilled worker is familiar. If, for example, integration into the DSB recognition site of the I-CpaI enzyme at the position identified by ^ (cggtcct^aaggtagcgaaattc) is desired, the mutated, adapted IGS can, for exam- 40 ple, have the following sequence: 5'-gggacc-3'.

In group II introns, which are mobile, further activities in addition to maturase are frequently encoded in the protein moiety of the ribonucleoprotein complex. However, these are not 45 necessarily required for the method described and can therefore be deleted. Indeed, deletion is preferred since it makes the construct in question smaller and easier to handle. The skilled

## 46

worker is familiar with a variety of options for removing such activities from the protein moiety. For example, this can be effected by generating a synthetic gene which comprises only the desired regions, or by suitable PCR methods.

5

Self-splicing group II introns have a conserved structure and generally consist of 6 different domains. Domain I comprises the exon binding sites (EBS1 and EBS2) which, during the splicing procedure, interact with the exon located 5' from the intron. In addition, an interaction between the "δ region" (located immediately 5' of EBS1) and the "δ' region" at the 3' exon takes place (Lambowitz AM & Belfort M (1993) Annu Rev Biochem 62:587-622; Michel F & Ferat JL (1995) Annu Rev Biochem 64:435-461). These sequences can be adapted by techniques with which the skilled worker is familiar, such as synthetic generation of the introns or suitable PCR methods, in each case in such a way that correct choice of the splicing sites at the insert site chosen in the DSB recognition region is ensured. This is done in such a way that the regions mentioned are modified so that base pairings with the corresponding sequences upstream (intron binding sequences, IBS) and downstream (δ') of the artificial insertion sequence can be undergone. If, for example, cggtcctaaggt^agcgaaattc is chosen as insertion site (^) for the Ll.LtrB intron in the I-CpaI recognition region, the δ region and the EBS1 region can, for example, adopt the sequence TCGCTACCTTAG (natural sequence: TTATGGTTGTG), and EBS2 for example the sequence GACCG (natural sequence: ATGTG). If the Arabidopsis thaliana trnK intron is selected, the δ region and the EBS1 region can, for example, adopt the sequence CGCTACCTTAGG (natural sequence: AATGTTAAAA), assuming the same insertion site as indicated for the Ll.LtrB intron.

If the DSB recognition sequence takes the form of a natural, endogenous recognition sequence of a homing endonuclease, a selected intron is preferably inserted at the site of the DSB recognition region at which the intron belonging to the homing endonuclease in question can also be found naturally.

The artificial insertion site of an intron in the DSB recognition site is preferably chosen such that 5' and 3' of the intron inserted as many bases as possible correspond to those of the natural insertion site of the intron in question and that the DSB recognition sequence is no longer functional after insertion of the intron. Very especially preferably, the nucleotide located in each case immediately upstream or downstream of the insert site of the intron corresponds to that at the natural insertion site.

## 47

In an especially preferred embodiment, the intron is flanked by homology sequences in order to make possible a directed insertion. Here, the homology sequences are — as described above — homologous to the sequences flanking the DSB recognition sequence and thus make possible an accurate insertion.

The invention therefore furthermore also relates to DNA constructs comprising at least one nucleic acid and intron sequence elements which are capable of ensuring, in a ribonucleic acid sequence derived from said DNA construct, the deletion of the ribonucleic acid fragment encoding said nucleic acid sequence, where said nucleic acid sequence is heterologous with regard to said intron sequence elements.

In a preferred embodiment, the nucleic acid sequence is flanked at least by a splice acceptor sequence and a splice donor sequence.

In a further embodiment, the DNA construct comprises, at the 5' end, sequences H1 and H2, respectively, which have sufficient length and homology with plastid sequences H1' and H2', respectively, to ensure homologous recombination between H1 and H1', and H2 and H2', respectively, and thus insertion of the H1- and H2-flanked sequence into the plastome.

The invention furthermore relates to a transgenic plastidic DNA comprising at least one nucleic acid sequence and intron sequence elements which are capable of ensuring, in a ribonucleic acid sequence derived from said transgenic plastidic DNA, the deletion of said ribonucleic acid fragment encoding said nucleic acid sequence, where said nucleic acid sequence is heterologous with regard to said intron sequence elements. In a preferred embodiment, the nucleic acid sequence is flanked by at least one splice acceptor sequence and one splice donor sequence.

To construct a transformation vector, the insertion sequence or the transformation construct can be cloned into a standard vector such as pBluescript or pUC18. In a preferred embodiment, the insertion sequence or the transformation construct is applied as a linear or linearized DNA molecule.

Preferably, only the portion of transformation vector which comprises the insertion sequence or the transformation construct with, if appropriate, homology sequences, selection marker and/or the expression cassette for the DSB1 enzyme is applied. If all or some of the homology sequences are dispensed with, the linearized DNA molecule is preferably obtained by digestion with restriction

## 48

endonucleases which generate single-stranded DNA overhangs at one or at both ends which are compatible with those generated by the DSBI enzyme in the plastidic DNA.

5 In a preferred embodiment, the transformation vector can comprise elements (for example a plastidic ORI (*origin of replication*)), which make it possible for the vector autonomously to replicate in the plastid or stably to exist in the plastids as extrachromosomal DNA molecule, before being integrated into the  
10 plastidic DNA. Such methods are known to the skilled worker (US 5,693,507; US 5,932,479; WO 99/10513). This method is preferred since it increases the copy number of the insertion sequences which is available for integration in the plastid.

15 One of the above-described constructs can be introduced into the plastids of a suitable master plant using one of the methods described. Microinjection is preferred, particle bombardment is particularly preferred.

20 Cloning, expression, selection and transformation methods

"Expression cassette" means -- for example regarding the expression cassette for the DSBI enzyme -- those constructions in which the DNA to be expressed is in operable linkage with at  
25 least one genetic control element which makes possible or regulates its expression (i.e. transcription and/or translation). In this context, expression can be, for example, stable or transient, constitutive or inducible. A variety of direct methods (for example transfection, particle bombardment, microinjection)  
30 or indirect methods (for example agrobacterial infection, viral infection) stated hereinbelow is available to the skilled worker for the introduction, and these methods will be stated hereinbelow.

35 An operable linkage is generally understood as meaning an arrangement in which a genetic control sequence can exert its function with regard to a nucleic acid sequence, for example encoding a DSBI enzyme. Function, in this context, can mean for example the control of the expression, i.e. transcription and/or  
40 translation, of the nucleic acid sequence, for example encoding a DSBI enzyme. Control, in this context, comprises for example initiating, increasing, governing or suppressing the expression, i.e. transcription and, if appropriate, translation. Governing, in turn, can be tissue and/or timing-specific. It may also be  
45 inducible, for example by certain chemicals, stress, pathogens and the like.

## 49

An operable linkage is understood as meaning for example the sequential arrangement of a promoter, of the nucleic acid sequence to be expressed - for example encoding a DSB1 enzyme - and, if appropriate, further regulatory elements such as, for example, a terminator in such a way that each of the regulatory elements can fulfill its function upon expression of the nucleic acid sequence, for example encoding a DSB1 enzyme. In this context, operable linkage need not necessarily exist on the transformation constructs themselves. Operable linkage can also result as a consequence of the insertion into the nuclear or plastidic DNA, where the regulatory elements are already present in the nuclear or plastidic DNA. In this respect, the regulatory elements can be naturally present or else introduced in a preceding step, for example when introducing an artificial DSB recognition sequence.

Direct linkage in the chemical sense is not necessarily required in this context. Genetic control sequences such as, for example, enhancer sequences, can also exert their function on the target sequence from positions which are further away, or indeed from other DNA molecules. Preferred arrangements are those in which the nucleic acid sequence to be expressed - for example encoding a DSB1 enzyme - is positioned behind a sequence which acts as promoter, so that the two sequences are bonded covalently with one another. Preferably, the distance between the promoter sequence and the nucleic acid sequence - for example encoding a DSB1 enzyme - is less than 200 base pairs, especially preferably less than 100 base pairs, very especially preferably less than 50 base pairs.

The skilled worker is familiar with a variety of routes to obtain one of the transformation constructs according to the invention, vectors comprising them or one of the expression cassettes. They can be prepared by means of customary recombination and cloning techniques as are described for example in Maniatis T, Fritsch EF and Sambrook J, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989) and in Silhavy TJ, Berman ML and Enquist LW, Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984) and in Ausubel FM et al., Current Protocols in Molecular Biology, Greene Publishing Assoc. and Wiley Interscience (1987). The direct fusion of a nucleic acid sequence which acts as promoter and a nucleotide sequence to be expressed - for example encoding a DSB1 enzyme - is preferred.

## 50

The term "genetic control sequences" is to be understood in the broad sense and refers to all those sequences which influence the generation or the function of an expression cassette or transformation vector. Genetic control sequences ensure  
5 transcription and, if appropriate, translation in the nucleus (or cytoplasm) or plastids. Preferably, the expression cassettes according to the invention comprise a promoter 5' upstream of the respective nucleic acid sequence to be expressed and a terminator  
10 sequence as additional genetic control sequence 3' downstream, and, if appropriate, further customary regulatory elements, in each case in operable linkage with the nucleic acid sequence to be expressed.

Genetic control sequences are described, for example, by  
15 "Goeddel; Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990)" or "Gruber and Crosby, in: Methods in Plant Molecular Biology and Biotechnology, CRC Press, Boca Raton, Florida, eds.: Glick and Thompson, Chapter 7, 89-108" and the references cited therein.

20 Examples of such control sequences are sequences to which the inductors or repressors bind and thus regulate the expression of nucleic acid. The natural regulation of these sequences may still be present before the actual structural genes, in addition to  
25 these novel control sequences or instead of these sequences, and, if appropriate, can have been genetically modified so that the natural regulation has been switched off and gene expression enhanced. However, the expression cassette can also be simpler in structure, that is to say no additional regulatory signals are  
30 inserted before the abovementioned genes and the natural promoter together with its regulation is not removed. Instead, the natural control sequence is mutated in such a way that regulation no longer takes place and gene expression is enhanced. These modified promoters can also be placed by themselves before the  
35 natural genes in order to increase the activity.

Depending on the host organism or the starting organism described in greater detail hereinbelow, which is converted into a genetically modified or transgenic organism by the introduction  
40 of the expression cassettes or vectors, different control sequences are suitable.

Promoters which are suitable for nuclear expression (for example of a viral/bacteriophage RNA polymerase or of a DSBII enzyme with  
45 plastidic transit peptide) are, in principle, all those which are

## 51

capable of governing the expression of genes, in particular foreign genes, in plants.

- Suitable promoters are those which make possible constitutive expression in plants (Benfey et al. (1989) EMBO J. 8:2195-2202). In particular, a plant promoter or a promoter derived from a plant virus is used by preference. Especially preferred is the promoter of the cauliflower mosaic virus 35S transcript (Franck et al. (1980) Cell 21:285-294; Odell et al. (1985) Nature 313:810-812; Shewmaker et al. (1985) Virology 140:281-288; Gardner et al. 1986, Plant Mol. Biol. 6, 221-228) or the 19S CaMV promoter (US 5,352,605 and WO 84/02913). As is known, this promoter comprises different recognition sequences for transcriptional effectors which, in their totality, lead to largely permanent and constitutive expression of the gene introduced (Benfey et al. (1989) EMBO J 8:2195-2202). A further suitable constitutive promoter is the "Rubisco small subunit (SSU)" promoter (US 4,962,028). A further example of a suitable promoter is the leguminB promoter (GenBank Acc. No.: X03677). Examples of further preferred constitutive promoters are the Agrobacterium nopaline synthase promoter, the TR dual promoter, the Agrobacterium OCS (octopine synthase) promoter, the ubiquitin promoter (Holtorf S et al. (1995) Plant Mol Biol 29:637-649), the promoters of the vacuolar ATPase subunits, the FBPaseP promoter (WO 98/18940) or the promoter of a proline-rich protein from wheat (WO 91/13991). Other suitable constitutive promoters which are preferred for the purposes of the present invention are the Super promoter (Ni M et al. (1995) Plant J 7:661-676; US 5,955,646) and the nitrilase-1 promoter of the Arabidopsis nit1 gene (GenBank Acc. No.: Y07648.2, nucleotides 2456 to 4340; Hillebrand H et al. (1998) Plant Mol Biol 36 (1):89-99; Hillebrand H et al. (1996) Gene 170(2):197-200).

- Promoters which are preferred are inducible promoters, especially preferably chemically inducible promoters (Aoyama T and Chua NH (1997) Plant J 11:605-612; Caddick MX et al. (1998) Nat. Biotechnol 16:177-180; Review: Gatz (1997) Annu Rev Plant Physiol Plant Mol Biol 48:89-108) by means of which expression can be controlled at a particular point in time. Examples which may be mentioned are the PRP1 promoter (Ward et al. (1993) Plant Mol Biol 22:361-366), a salicylic-acid-inducible promoter (WO 95/19443), a benzenesulfonamide-inducible promoter (EP-A-0388186), a tetracyclin-inducible promoter (Gatz et al. (1992) Plant J 2:397-404), an abscisic-acid-inducible promoter (EP-A 335 528), an ethanol-inducible promoter (Salter MG et al. (1998) Plant J. 16:127-132), the heavy-metal-inducible metallothionein I promoter (Amini S et al. (1986) Mol Cell Biol



## 52

- 6:2305-2316), the steroid-inducible MMTV LTR promoter (Izant JG et al. (1985) Science 229:345-352) and a cyclohexanone-inducible promoter (WO 93/21334). Especially preferred is the inducible expression of a PLS/DSBI enzyme fusion protein in the nucleus.
- 5 Inducible promoters also comprise those which are capable of regulation by certain repressor proteins (for example tet, lac). Such repressor proteins can be translocated into the plastids in fusion with PLS, where they regulate the expression of certain genes under the control of suitable promoters. In the plastids,
- 10 the repressor binds to an artificial repressor binding site which has been introduced into the plastome and can thus repress the expression of the downstream gene (cf. WO 95/25787). In this manner it is possible, for example, to induce the expression of a plastid-encoded DSBI enzyme when required, or to repress it until
- 15 the point in time at which expression is desired.

- Other promoters which are preferred are those which are induced by biotic or abiotic stress such as, for example, the pathogen-inducible promoter of the PRP1 gene (Ward et al.,
- 20 Plant Mol Biol 1993, 22: 361-366), the heat-inducible tomato hsp70 promoter or hsp80 promoter (US 5,187,267), the chill-inducible potato alpha-amylase promoter (WO 96/12814) or the wounding-induced pinII promoter (EP-A 0 375 091).
- 25 In an especially preferred embodiment, the nucleic acid which encodes the DSBI enzyme is, above all, expressed under the control of an inducible promoter. A controlled expression capable of being governed is thus obtained, and any problems caused by expressing a DSBI enzyme constitutively are avoided.
- 30 Advantageous control sequences for the expression cassettes or vectors according to the invention comprise viral, bacteriophage or bacterial promoters such as cos, tac, trp, tet, phoA, tat, lpp, lac, lacIq, T7, T5, T3, gal, trc, ara, SP6,  $\lambda$ -PR or  $\lambda$ -PL
- 35 promoter. They are preferably employed in combination with the expression of the respective, corresponding RNA polymerase.

- The expression in plastids can be effected using plastid promoters and/or transcription regulation elements. Examples
- 40 which may be mentioned, but not by way of limitation, are the RNA polymerase promoter (WO 97/06250) or the promoters described in WO 00/07431, US 5,877,402, WO 97/06250, WO 98/55595, WO 99/46394, WO 01/42441 and WO 01/07590. The rpo B promoter element, the atpB promoter element, the clpP promoter element (see also WO
- 45 99/46394) or the 16S rDNA promoter element should be mentioned. In this context, the promoter can also have a polycistronic "operon" assigned to it (EP-A 1 076 095; WO 00/20611). Systems in

## 53

which a nonplant (for example viral) RNA polymerase is imported into the plastid using plastidic transit peptides and specifically induces, in the plastid, the expression transgenic sequences which are under the control of the RNA polymerase recognition sequences and have previously been inserted into the plastidic DNA have also been described (WO 95/16783; US 5,925,806; US 5,575,198).

In addition to the abovementioned promoters, the following can also be preferably used:

- a) the PrbcL promoter (SEQ ID NO: 44)
- b) the Prps16 promoter (SEQ ID NO: 50)
- c) the Prn16 promoter (SEQ ID NO: 46)

In an especially preferred embodiment, NEP promoters are employed. These are promoters which are functional in plastids and are recognized by the nuclear-encoded plastidic RNA polymerases (NEP). The following are preferred: Prn-62; Pycf2-1577; PatpB-289; Prps2-152; Prps16-107; Pycf1-41; PatpI-207; PclpP-511; PclpP-173 and PaccD-129 (WO 97/06250; Hajdukiewicz PTJ et al. (1997) EMBO J 16:4041-4048).

The following are especially preferred:

- a) the PaccD-129 promoter of the tobacco accD gene (WO 97/06250; SEQ ID NO: 47)
- b) the PclpP-53 promoter of the clpP gene as highly active NEP promoter in chloroplasts (WO 97/06250; SEQ ID NO: 48)
- c) the Prn-62 promoter of the rrn gene (SEQ ID NO: 49)
- d) the Prps16-107 promoter of the rps16 gene (SEQ ID NO: 45)
- e) the PatpB/E-290 promoter of the tobacco atpB/E gene (Kapoor S et al. (1997) Plant J 11:327-337) (SEQ ID NO: 51)
- f) the PrpoB-345 promoter of the rpoB gene (Liere K & Maliga P (1999) EMBO J 18: 249-257) (SEQ ID NO: 52)

In general, all those promoters which belong to class III (Hajdukiewicz PTJ et al. (1997) EMBO J 16:4041-4048) and all fragments of the class II promoters which control the initiation of transcription by NEP can be utilized in this preferred

## 54

embodiment. Such promoters or promoter moieties are not particularly highly conserved. ATAGAATAAA is given as consensus near the transcription initiation site of NEP promoters (Hajdukiewicz PTJ et al (1997) EMBO J 16:4041-4048).

5

Normally, genes are surrounded by regulatory sequences which originate from the plastids of the organism to be transformed. Thus, sequence duplications which can lead to instabilities owing to spontaneous, intrachromosomal homologous recombination events are generated (Heifetz PB (2000) Biochimie 82(6-7):655-666). To overcome this problem, it has been proposed to utilize heterologous regulatory sequences or to exploit regulatory units which already exist endogenously in the plastidic genome (WO 99/46394; WO 01/42441). A reduction of the homology by mutagenesis of the endogenous promoter sequence has also been described (WO 01/07590).

In principle, all natural promoters together with their regulatory sequences, such as those mentioned above, can be used for the method according to the invention. Especially preferred promoters are those which have been isolated from prokaryotes. Very especially preferred are promoters isolated from *Synechocystis* or *E.coli*. In addition, synthetic promoters such as, for example, a synthetic promoter derived from the *E. coli* consensus sequence for  $\sigma 70$  promoters

5'-TTGACA N<sub>16-19</sub> TATAAT N<sub>3</sub> CAT -3',

where N represents any nucleotide (that is to say A, G, C or T) can additionally also be used advantageously. It is obvious to the skilled worker that individual or few base substitutions in the conserved regions stated are also possible without destroying the function of the promoter. The variable design of these synthetic promoters by using a variety of sequential sequences makes it possible to generate a multiplicity of promoters which lack extensive homologies, which increases the stability of the expression cassettes in the plastome in particular in the event that several promoters are required. The following, particularly preferred promoter sequences, which are derived from the abovementioned consensus sequence, may be mentioned by way of example, but not by limitation:

- a) 5'-TTGACATTCACCTCTTCAATTATCTATAATGATACA-3' (SEQ ID NO: 53)
- b) 5'-TTGACAATTTTCCTCTGAATTATATAATTAACAT-3' (SEQ ID NO: 72)

45

## 55

It is obvious to the skilled worker that these synthetic promoters can control the expression of any genes. For example, they can be utilized for driving the expression of a selection marker, also in order to be able to select under regenerative conditions for transplastomic plants with the aid of said selection system. Examples of selection markers are enumerated further below. In addition, such synthetic promoters can be linked with any gene, for example with genes encoding antibodies, antigens or enzymes. Preferably, the expression cassettes consisting of such promoters also comprise 5'-untranslated regions (or ribosome binding sites) or 3'-noncoding regions which are detailed hereinbelow.

The invention furthermore relates to expression cassettes comprising a nucleic acid sequence encoding a DSBI enzyme under the control of a promoter which is functional in plant plastids, for example one of the above-described promoters. The expression cassette can comprise further elements such as, for example, transcription terminators and/or selection markers.

Genetic control sequences also comprise further promoters, promoter elements or minimal promoters which are capable of modifying the expression-controlling properties. Genetic control sequences furthermore also comprise the 5'-untranslated region (5'-UTR) or the noncoding 3' region (3'-UTR) of genes (Eibl C (1999) Plant J 19: 1-13). It has been demonstrated that these can exert significant functions in regulating the gene expression in plastids of higher plants. In the nucleus, too, genetic control elements such as 5'-UTR, introns or 3'-UTR, can exert a function in gene expression. Thus, for example, it has been demonstrated that 5'-untranslated sequences can enhance the transient expression of heterologous genes. They can furthermore promote tissue specificity (Rouster J et al., Plant J. 1998, 15: 435-440.).

5'-UTRs and 3'-UTRs which are preferably employed in plastids are:

- a) 5'psbA (from tobacco) (SEQ ID NO: 54)
- b) 5'rbcl including 5' portions from the coding region of the rbcl gene (from tobacco) (SEQ ID NO: 55); the sequence described as SEQ ID NO: 55 has been mutated in comparison with the native sequence in order to introduce a PstI and an NcoI cleavage site.

## 56

c) 5'rbCLs (SEQ ID NO: 56); the sequence described by SEQ ID NO: 56 has been mutated in comparison with the native sequence in order to introduce a *PagI* cleavage site.

5 d) 3'psbA-1 from *Synechocystis* (SEQ ID NO: 57)

e) 3'psbA from tobacco (SEQ ID NO: 58)

f) 3'rbCL from tobacco (SEQ ID NO: 59)

10

Genetic control sequences, especially for expression in plastids, also comprise in particular ribosome binding sequences for initiating translation. They are usually present in the 5'-UTRs. This is especially preferred when suitable sequences are not  
15 provided by the nucleic acid sequence to be expressed or when such sequences are compatible with the expression system. Especially preferred is the use of a synthetic ribosome binding site (RBS) with the sequence 5'-GGAGG(N)<sub>3-10</sub>ATG-3', preferably 5'-GGAGG(N)<sub>5</sub>ATG-3' (SEQ ID NO: 60), particularly preferably  
20 5'-GGAGGATCTCATG-3' (SEQ ID NO: 61).

The expression cassette can advantageously comprise one or more what are known as enhancer sequences in operable linkage with the promoter; these enhancer sequences make possible an enhanced  
25 transgenic expression of the nucleic acid sequence. Additional advantageous sequences, such as further regulatory elements or terminators, may also be inserted at the 3' end of the nucleic acid sequences to be expressed recombinantly. One or more copies of the nucleic acid sequences to be expressed recombinantly may  
30 be present in the gene construct.

It is furthermore possible to insert, after the start codon, what is known as a downstream box, which enhances expression in general (translation enhancer WO 00/07431; WO 01/21782).

35

Polyadenylation signals which are suitable as genetic control sequences, above all in the transformation of the nucleus, are plant polyadenylation signals, preferably those which correspond essentially to T-DNA polyadenylation signals from *Agrobacterium*  
40 *tumefaciens*, in particular of gene 3 of the T-DNA (octopine synthase) of the Ti plasmid pTiACHS (Gielen et al., EMBO J. 3 (1984), 835 et seq.) or functional equivalents thereof. Examples of especially suitable terminator sequences are the OCS (octopine synthase) terminator and the NOS (nopaline synthase) terminator.

45

## 57

The transformation vectors and insertion sequences according to the invention may comprise further nucleic acid sequences. Such nucleic acid sequences can preferably constitute expression cassettes. The following may be mentioned by way of example of the DNA sequences to be expressed in the expression constructs, but not by limitation:

## 1. Selection markers

- 10 "Selection markers" means all those nucleic acid or protein sequences whose expression (i.e. transcription and, if appropriate, translation) confers a phenotype to a cell, tissue or organism which differs from that of an untransformed cell, tissue or organism. Selection markers comprises for example those
- 15 nucleic acids or protein sequences whose expression confers an advantage (positive selection marker) or disadvantage (negative selection marker) on a cell, tissue or organism in comparison with cells which do not express this nucleic acid or protein. For example, positive selection markers act by detoxifying a
- 20 substance which has an inhibitory effect on the cell (for example resistance to antibiotics/herbicides), or by forming a substance which makes possible improved regeneration or enhanced growth of the plant under the selected conditions (for example nutritive markers, hormone-producing markers such as ipt; see hereinbelow).
- 25 Another form of positive selection markers comprises mutated proteins or RNAs which are insensitive to a selective agent (for example 16S rRNA mutants, which are insensitive to spectinomycin). Negative selection markers act for example by catalyzing the formation of a toxic substance in the transformed
- 30 cells (for example the codA gene). Moreover, selection marker can also comprise reporter proteins as long as they are suitable for differentiating transformed from untransformed cells, tissues or organs (for example by coloration or another detectable phenotype).

35

The following selection markers may be mentioned by way of example, but not by limitation:

## 1.1 Positive selection markers:

40

- The selectable marker introduced into the nucleus or the plastids together with the expression cassette confers resistance to a biocide (for example a herbicide such as phosphinothricin, glyphosate or bromoxynil), a metabolic inhibitor such as
- 45 2-deoxyglucose-6-phosphate (WO 98/45456) or an antibiotic such as, for example, tetracyclins, ampicillin, kanamycin, G 418, neomycin, bleomycin or hygromycin, to the successfully

## 58

transformed cells. The selection marker permits the selection of the transformed cells from untransformed cells (McCormick et al., Plant Cell Reports 5 (1986), 81-84; Dix PJ & Kavanagh TA (1995) Euphytica 85: 29-34).

5

Especially preferred selection markers are those which confer resistance to herbicides. Selection markers which may be mentioned by way of example are:

- 10 - DNA sequences which encode phosphinothricin acetyltransferases (PAT), which acetylate the free amino group of the glutamine synthase inhibitor phosphinothricin (PPT) and thus detoxify the PPT (de Block et al. 1987, EMBO J. 6, 2513-2518) (also referred to as Bialophos® resistance gene (bar)). The bar gene encoding a phosphinothricin acetyltransferase (PAT) can be isolated for example from Streptomyces hygroscopicus or S. viridochromogenes. Such sequences are known to the skilled worker (from Streptomyces hygroscopicus GenBank Acc. No.: X17220 and X05822, from Streptomyces viridochromogenes GenBank Acc. No.: M 22827 and X65195; US 5,489,520). Synthetic genes are further described for example for expression in plastids. A synthetic PAT gene is described in Becker et al. (1994) The Plant J. 5:299-307. The genes confer resistance to the herbicide bialaphos or glufosinate and are widely used markers in transgenic plants (Vickers JE et al. (1996). Plant Mol Biol Reporter 14:363-368; Thompson CJ et al. (1987) EMBO J 6:2519-2523).
- 5-Enolpyruvylshikimate-3-phosphate synthase genes (EPSP synthase genes), which confer resistance to glyphosate (N-(phosphonomethyl)glycine). The nonselective herbicide glyphosate has 5-enolpyruvyl-3-phosphoshikimate synthase (EPSPS) as molecular target. This enzyme has a key function in the biosynthesis of aromatic amino acids in microbes and plants, but not in mammals (Steinrücken HC et al. (1980) Biochem. Biophys. Res. Commun. 94:1207-1212; Levin JG and Sprinson DB (1964) J. Biol. Chem. 239: 1142-1150; Cole DJ (1985) Mode of action of glyphosate a literature analysis, p. 48-74. In: Grossbard E and Atkinson D (eds.). The herbicide glyphosate. Butterworths, Boston.). Glyphosate-tolerant EPSPS variants are preferably used as selection markers (Padgett SR et al. (1996). New weed control opportunities: development of soybeans with a Roundup Ready™ gene. In: Herbicide Resistant Crops (Duke, S.O., ed.), pp. 53-84. CRC Press, Boca Raton, FL; Saroha MK and Malik VS (1998) J Plant Biochemistry and Biotechnology 7:65-72). The EPSPS gene of Agrobacterium sp. strain CP4 has a natural

## 59

- tolerance to glyphosate which can be transferred to corresponding transgenic plants. The CP4 EPSPS gene has been cloned from *Agrobacterium* sp. strain CP4 (Padgett et al. (1995) Crop Science 35(5):1451-1461). Sequences
- 5 of 5-enolpyruvylshikimate-3-phosphate synthases which are glyphosate-tolerant, such as, for example, those described in US 5,510,471; US 5,776,760; US 5,864,425; US 5,633,435; US 5,627,061; US 5,463,175; EP 0 218 571, are described in the patents and also deposited in GenBank. Further sequences
- 10 are described under GenBank Accession X63374. The *aroA* gene (M10947 *S. typhimurium* *aroA* locus 5-enolpyruvylshikimate-3-phosphate synthase (*aroA* protein) gene) is furthermore preferred.
- 15 - the *gox* gene (glyphosate oxidoreductase), which encodes the Glyphosat® degrading enzyme. GOX (for example the glyphosate oxidoreductase from *Achromobacter* sp.) catalyzes the cleavage of a C-N bond in glyphosate, which is thus converted into aminomethylphosphonic acid (AMPA) and glyoxylate. GOX can
- 20 thereby confer resistance to glyphosate (Padgett et al. (1996) J Nutr. 1996 Mar; 126(3):702-16; Shah D et al. (1986) Science 233: 478-481).
- the *deh* gene (encoding a dehalogenase which inactivates Dalapon®), (GenBank Acc. No.: AX022822, AX022820 and WO 99/27116)
- 25
- *bxn* genes, which encode bromoxynil-degrading nitrilase enzymes, for example the *Klebsiella ozanenae* nitrilase.
- 30 Sequences can be found in GenBank for example under the Acc. No: E01313 (DNA encoding bromoxynil specific nitrilase) and J03196 (*K. pneumoniae* bromoxynil-specific nitrilase (*bxn*) gene, complete cds).
- 35 - Neomycin phosphotransferases confer resistance to antibiotics (aminoglycosides) such as neomycin, G418, hygromycin, paromomycin or kanamycin, by reducing their inhibitory action by means of a phosphorylation reaction. Especially preferred is the *nptII* gene. Sequences can be obtained from GenBank
- 40 (AF080390 minitransposon mTn5-GNm; AF080389 minitransposon mTn5-Nm, complete sequence). Moreover, the gene is already a component in a large number of expression vectors and can be isolated from them using methods with which the skilled worker is familiar (such as, for example, polymerase chain
- 45 reaction) (AF234316 pCAMBIA-2301; AF234315 pCAMBIA-2300, AF234314 pCAMBIA-2201). The NPTII gene encodes an aminoglycoside 3'-O-phosphotransferase from *E. coli*, Tn5



## 60

- (GenBank Acc. No: U00004 position 1401-2300; Beck et al. (1982) Gene 19 327-336). Moreover, the *Acinetobacter baumannii* aphA-6 gene, which encodes an aminoglycoside phosphotransferase, may also be utilized as selection marker (Huang et al. (2002) Mol Genet Genomics 268:19-27)
- 5
- the DOG<sup>R1</sup> gene. The gene DOG<sup>R1</sup> was isolated from the yeast *Saccharomyces cerevisiae* (EP 0 807 836). It encodes a 2-deoxyglucose-6-phosphate phosphatase, which confers resistance to 2-DOG (Randez-Gil et al. 1995, Yeast 11, 1233-1240; Sanz et al. (1994) Yeast 10:1195-1202, sequence: GenBank Acc. No.: NC001140 chromosome VIII, *Saccharomyces cerevisiae* position 194799-194056).
- 10
- 15 - Sulfonylurea- and imidazolinone-inactivating acetolactate synthases, which confer resistance to imidazolinone/sulfonylurea herbicides. Examples which may be mentioned of imidazolinone herbicides are the active substances imazamethabenz-methyl, imazamox, imazapyr, imazaquin and imazethapyr. Examples of sulfonylurea herbicides which may be mentioned are amidosulfuron, azimsulfuron, chlorimuronethyl, chlorsulfuron, cinosulfuron, imazosulfuron, oxasulfuron, prosulfuron, rimsulfuron, sulfosulfuron. The skilled worker is familiar with a large number of further active substances from the abovementioned classes. Nucleic acid sequences such as, for example, the sequence for the *Arabidopsis thaliana* Csr 1.2 Gen (EC 4.1.3.18) which has been deposited under the GenBank Acc No.: X51514, are suitable (Sathasivan K et al. (1990) Nucleic Acids Res. 18(8):2188). Acetolactate synthases, which confer resistance to imidazolinone herbicides, are furthermore described under the GenBank Acc. Nos.:
- 20
- 25
- 30
- 35
- a) AB049823 *Oryza sativa* ALS mRNA for acetolactate synthase, complete cds, herbicide resistant biotype
- b) AF094326 *Bassia scoparia* herbicide resistant acetolactate synthase precursor (ALS) gene, complete cds
- 40
- c) X07645 Tobacco acetolactate synthase gene, ALS SuRB (EC 4.1.3.18)
- d) X07644 Tobacco acetolactate synthase gene, ALS SuRA (EC 4.1.3.18)
- 45
- e) A19547 Synthetic nucleotide mutant acetolactate synthase

## 61

- f) A19546 Synthetic nucleotide mutant acetolactate synthase
- g) A19545 Synthetic nucleotide mutant acetolactate synthase
- 5 h) I05376 Sequence 5 from patent EP 0257993
- i) I05373 Sequence 2 from patent EP 0257993
- j) AL133315
- 10 - Hygromycin phosphotransferases (X74325 *P. pseudomallei* gene for hygromycin phosphotransferase) which confer resistance to the antibiotic hygromycin. The gene is a component of a large number of expression vectors and can be isolated from them
- 15 using methods with which the skilled worker is familiar (such as, for example, polymerase chain reaction) (AF294981 pINDEX4; AF234301 pCAMBIA-1380; AF234300 pCAMBIA-1304; AF234299 pCAMBIA-1303; AF234298 pCAMBIA-1302; AF354046 pCAMBIA-1305.; AF354045 pCAMBIA-1305.1)
- 20 - genes for resistance to
- a) chloramphenicol (chloramphenicol acetyltransferase),
- 25 b) tetracyclin, various resistance genes have been described, for example X65876 *S. ordonez* genes class D tetA and tetR for tetracyclin resistance and repressor proteins X51366 *Bacillus cereus* plasmid pBC16 tetracyclin resistance gene. Moreover, the gene is already a
- 30 component of a large number of expression vectors and can be isolated therefrom using methods known to the skilled worker (such as, for example polymerase chain reaction)
- 35 c) Streptomycin; various resistance genes have been described, for example with the GenBank Acc. No.:AJ278607 *Corynebacterium acetoacidophilum* ant gene for streptomycin adenylyltransferase.
- 40 d) Zeocin; the corresponding resistance gene is a component of a large number of cloning vectors (for example L36849 cloning vector pZEO) and can be isolated from these using methods known to the skilled worker (such as, for example, polymerase chain reaction).
- 45 e) Ampicillin ( $\beta$ -lactamase gene; Datta N, Richmond MH. (1966) *Biochem J.* 98(1):204-9; Heffron F et al. (1975) *J. Bacteriol* 122: 250-256; the Amp gene was first cloned

## 62

for generating the *E. coli* vector pBR322; Bolivar F et al. (1977) Gene 2:95-114). The sequence is a component of a large number of cloning vectors and can be isolated from them using methods known to the skilled worker (such as, for example, polymerase chain reaction).

- Genes such as the isopentenyl transferase from *Agrobacterium tumefaciens* (strain:PO22) (Genbank Acc. No.: AB025109). The *ipt* gene is a key enzyme of cytokinin biosynthesis. Its overexpression facilitates the regeneration of plants (for example selection on cytokinin-free medium). The method for utilizing the *ipt* gene has been described (Ebinuma H et al. (2000) Proc Natl Acad Sci USA 94:2117-2121; Ebinuma, H et al. (2000) Selection of marker-free transgenic plants using the oncogenes (*ipt*, *rol* A, B, C) of *Agrobacterium* as selectable markers, In Molecular Biology of Woody Plants. Kluwer Academic Publishers).

Various other positive selection markers which confer a growth-related advantage to the transformed plants over the nontransformed plants, and methods for their use, have been described, inter alia, in EP-A 0 601 092. Examples which may be mentioned are  $\beta$ -glucuronidase (in conjunction with, for example, cytokinin glucuronide), mannose-6-phosphate isomerase (in conjunction with mannose), UDP-galactose 4-epimerase (in conjunction with, for example, galactose), with mannose-6-phosphate isomerase in conjunction with mannose being especially preferred.

Preferred for a selection marker which is functional in plastids are, in particular, those which confer resistance to spectinomycin, streptomycin, kanamycin, lincomycin, gentamycin, hygromycin, methotrexate, bleomycin, phleomycin, blasticidin, sulfonamide, phosphinothricin, chlorsulfuron, bromoxynil, glyphosate, 2,4-D, atrazine, 4-methyltryptophan, nitrate, S-aminoethyl-L-cysteine, lysine/threonine, aminoethyl-cysteine or betaine aldehyde. Especially preferred are the genes *aadA*, *nptII*, *BADH*, *FLARE-S* (a fusion of *aadA* and GFP, described by Khan MS & Maliga P, 1999 Nature Biotech 17: 910-915).

As selection marker which is functional in plastids, it is mainly the *aadA* gene which has been described (Svab Z and Maliga P (1993) Proc Natl Acad Sci USA 90:913-917). Also described are modified 16S rDNA, the *nptII* gene (kanamycin resistance) and the *bar* gene (phosphinothricin resistance). Owing to the preference given to the selection marker *aadA*, the latter is preferably recycled, i.e. deleted from the genome, or plastome, following

## 63

its use (Fischer N et al. (1996) Mol Gen Genet 251:373-380; Corneille S et al. (2001) Plant J 27:171-178), so that aadA can be reused as selection marker in further transformations of an already transplastomic plant. The betaine-aldehyde dehydrogenase (BADH) from spinach has been described as a further possible selection marker (Daniell H et al. (2001) Trends Plant Science 6:237-239; Daniell H et al. (2001) Curr Genet 39:109-116; WO 01/64023; WO 01/64024; WO 01/64850). Lethal agents such as, for example, glyphosate, can also be utilized in connection with correspondingly detoxifying or resistant enzymes (WO 01/81605).

Binding type markers may also be utilized. To utilize the DBS recognition sequence of the homing endonuclease I-CpaI in the gene of the 23S rRNA, which sequence is preferred as insertion site, at least the 3' end of the insertion sequence (preferably an artificial intron) is surrounded by homologous sequences of the target region. Thus, sequences of the 23S rDNA are incorporated into the transformation vector. Point mutations can be introduced at one position (position 2073 or 2074 of the tobacco 23S rRNA, sequence: AAAGACCCTATGAAG becomes sequence: GGAGACCCTATGAAG), which point mutations confer resistance to lincomycin to the ribosomes derived from a 23S rDNA which has been mutated thus (Cseplö A et al. (1988) Mol Gen Genet 214:295-299). Further point mutations comprise those in the tobacco 16S rRNA which confer resistance to spectinomycin (mutation underlined):

a) 5'-GGAAGGTGAGGATGC-3' (A in native sequence)

Other mutations confer resistance to streptomycin:

b) 5'-GAATGAAACTA-3' (C in native sequence)

## 1.2 Negative selection markers

Negative selection markers make possible for example the selection of organisms with successfully deleted sequences which comprise the marker gene (Koprek T et al. (1999) The Plant Journal 19(6):719-726). For example, sequences which encode selection markers or DSB1 enzymes can be deleted from the genome/plastome after successful application of the method according to the invention.

When carrying out a negative selection, for example a compound which otherwise has no disadvantageous effect on the plant is converted into a compound which is disadvantageous, for example owing to the negative selection marker introduced into the plant.

## 64

Genes which have a disadvantageous effect per se are furthermore also suitable, such as, for example, TK thymidine kinase (TK) and diphtheria toxin A fragment (DT-A), the *codA* gene product encoding a cytosine deaminase (Gleave AP et al. (1999) Plant Mol Biol 40(2):223-35; Perera RJ et al. (1993) Plant Mol Biol 23(4): 793-799; Stougaard J (1993) Plant J 3:755-761), the cytochrome P450 gene (Koprek et al. (1999) Plant J. 16:719-726), genes encoding a haloalkane dehalogenase (Naested H (1999) Plant J. 18:571-576), the *iaaH* gene (Sundaresan V et al. (1995) Genes & Development 9:1797-1810) or the *tms2* gene (Fedoroff NV & Smith DL (1993) Plant J 3:273-289).

The concentrations of the antibiotics, herbicides, biocides or toxins which are used in each case to carry out the selection must be adapted to the respective test conditions or organisms. Examples which may be mentioned in context with plants are: kanamycin (Km) 50 to 100 mg/l, hygromycin B 40 mg/l, phosphinothricin (Ppt) 6 to 20 mg/l, spectinomycin (Spec) 15 to 500 mg/l.

Furthermore, it is possible to express functional analogs of the abovementioned nucleic acids encoding selection markers. Functional analogs means, in the present context, all those sequences which have essentially the same function, i.e. which are capable of selecting transformed organisms. In this context, it is quite feasible that the functional analog differs with regard to other characteristics. For example, it can have a higher or lower activity, or else have further functionalities.

Functional analogs means furthermore sequences which encode fusion proteins consisting of one of the preferred selection markers and another protein, for example a further preferred selection marker, one of the reporter proteins mentioned hereinbelow or a PLS. By way of example, a fusion of the GFP (green fluorescent protein) and the *aadA* gene may be mentioned (Sidorov VA et al. (1999) Plant J 19:209-216).

## 2. Reporter genes

Reporter genes encode readily quantifiable proteins which, via their color or enzyme activity, allow an assessment of the transformation efficiency, the site or time of expression or the identification of transgenic plants. Very especially preferred in this context are genes encoding reporter proteins (see also Schenborn E, Groskreutz D. Mol Biotechnol. 1999; 13(1):29-44) such as

## 65

- "green fluorescent protein" (GFP) (Chui WL et al., Curr Biol 1996, 6:325-330; Leffel SM et al., Biotechniques. 23(5):912-8, 1997; Sheen et al. (1995) Plant Journal 8(5):777-784; Haseloff et al. (1997) Proc Natl Acad Sci USA 94(6):2122-2127; Reichel et al. (1996) Proc Natl Acad Sci USA 93(12):5888-5893; Tian et al. (1997) Plant Cell Rep 16:267-271; WO 97/41228).
- 5
- Chloramphenicol transferase,
- 10
- Luciferase (Millar et al., Plant Mol Biol Rep 1992 10:324-414; Ow et al. (1986) Science, 234:856-859); allows detection via bioluminescence.
- 15
- $\beta$ -Galactosidase, encodes an enzyme for which a variety of chromogenic substrates are available.
- $\beta$ -Glucuronidase (GUS) (Jefferson et al., EMBO J. 1987, 6, 3901-3907) or the uidA gene, which encodes an enzyme for a
- 20
- R-Locus gene product: protein which regulates the production of anthocyanin pigments (red coloration) in plant tissue and thus makes possible the direct analysis of the promoter
- 25
- activity without addition of further auxiliary substances or chromogenic substrates (Dellaporta et al., In: Chromosome Structure and Function: Impact of New Concepts, 18th Stadler Genetics Symposium, 11:263-282, 1988).
- 30
- $\beta$ -Lactamase (Sutcliffe (1978) Proc Natl Acad Sci USA 75:3737-3741), enzyme for a variety of chromogenic substrates (for example PADAC, a chromogenic cephalosporin).
- xylE gene product (Zukowsky et al. (1983) Proc Natl Acad Sci
- 35
- USA 80:1101-1105), catechol dioxygenase capable of converting chromogenic catechols.
- Alpha-amylase (Ikuta et al. (1990) Bio/technol. 8:241-242).
- 40
- Tyrosinase (Katz et al. (1983) J Gen Microbiol 129:2703-2714), enzyme which oxidizes tyrosine to DOPA and dopaquinone, which subsequently form melanin, which can be detected readily.

## 66

- Aequorin (Prasher et al. (1985) Biochem Biophys Res Commun 126(3):1259-1268), can be used in the calcium-sensitive bioluminescence detection.

5 The selection marker, or the reporter gene, is preferably encoded on the transformation construct, especially preferably on the insertion sequence. However, it can also be encoded on an independent transformation construct which is introduced into the nucleus or the plastids of a plant cell in the form of a  
10 cotransformation together with the transformation construct of interest.

The transformation vectors and insertion sequences according to the invention may comprise further functional elements. The  
15 concept of further functional elements is to be understood in the broad sense. Preferably, it is understood as meaning all those elements which influence the generation, multiplication, function, use or value of the insertion sequences, transformation constructs or transformation vectors used within the scope of the  
20 present invention. The following may be mentioned by way of example of further functional elements, but not by limitation:

- i. Replication origins (ORI) which make possible an amplification of the expression cassettes or vectors  
25 according to the invention in, for example, E. coli or else in plastids. Examples of E. coli ORIs which may be mentioned are the pBR322 ori, the P15A ori (Sambrook et al.: Molecular Cloning. A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989) or the colE1  
30 ORI, for example from pBLUESCRIPT. Plastidic ORIs are described in US 5,693,507, US 5,932,479 or WO 99/10513.
  - ii. Multiple cloning regions (MCRs) permit and facilitate the  
35 insertion of one or more nucleic acid sequences.
  - iii. Sequences which make possible the homologous recombination or insertion into the genome or plastome of a host organism.
  - iv. Elements, for example border sequences, which make possible  
40 the Agrobacterium-mediated transfer into plant cells for the transfer and integration into the plant genome, such as, for example, the right or left border of the T-DNA or the vir region.
- 45 An insertion sequence or an expression construct for a DSBI enzyme can be inserted advantageously using vectors into which these constructs, or cassettes, are inserted. Vectors can be

## 67

plasmids, cosmids, phages, viruses, retroviruses or else agrobacteria, by way of example.

In an advantageous embodiment, the expression cassette is  
5 inserted by means of plasmid vectors. Preferred vectors are those which make possible a stable integration of the expression cassette into the host genome or plastome.

The generation of a transformed organism or a transformed cell  
10 requires introducing the DNA in question into the host cell in question, or into the plastids thereof. A multiplicity of methods is available for this procedure, which is referred to as transformation (see also Keown et al. (1990) Methods in Enzymology 185:527-537). Thus, for example, the DNA can be  
15 introduced directly by microinjection, electroporation or by bombardment with DNA-coated microparticles. Also, the cell can be permeabilized chemically, for example with polyethylene glycol, so that the DNA can penetrate the cell by diffusion. Transformation can also be effected by fusion with other  
20 DNA-comprising units such as minicells, cells, lysosomes or liposomes. Others which must additionally be mentioned are transfection using calcium phosphate, DEAE dextran or cationic lipids, transduction, infection, the incubation of dry embryos in DNA-comprising solution, sonication, and the transformation of  
25 intact cells or tissue by microinjection or macroinjection into tissue or embryos, or tissue electroporation, or the vacuum infiltration of seeds. The skilled worker is familiar with such methods. In the case of injection or electroporation of DNA into plant cells, the plasmid used need not meet any particular  
30 requirements. Simple plasmids such as those from the pUC series can be used. If intact plants are to be regenerated from the transformed cells, the presence of an additional, selectable marker gene on the plasmid is useful. Methods for the regeneration of plants from plant tissues or plant cells have  
35 also been described.

There are several options for introducing DNA into the plastids. The only decisive aspect for the present invention is that the DNA is introduced into the plastids. However, the present  
40 invention is not limited to a specific method. Any method which permits the introduction of the DNA to be transformed into the plastids of a higher plant is suitable. The stable transformation of plastids is a method with which the skilled worker is familiar; it has been described for higher plants (inter alia by  
45 Svab Z and Maliga P (1993) Proc Natl Acad Sci USA 90(3):913-917). For example, the methods are based on transformation by means of the particle gun and insertion into the plastidic genome by



homologous recombination under selection pressure. Further methods are described in US 5,877,402. In EP-A 0 251 654, the DNA is introduced by *Agrobacterium tumefaciens* (see De Block M et al. (1985) EMBO J 4:1367-1372; Venkateswarlu K and Nazar RN (1991) 5 Bio/Technology 9:1103-1105). It has furthermore been demonstrated that DNA can be introduced into isolated chloroplasts by means of electroporation, thus bringing about transient expression (To KY et al. (1996) Plant J 10:737-743). Transformation by means of a direct DNA transfer into plastids of protoplasts, for example 10 using PEG (polyethylene glycol) is preferred (Koop HU et al. (1996) Planta 199:193-201; Kofer W et al. (1998) In Vitro Cell Dev Biol Plant 34:303-309; Dix PJ and Kavanagh TA (1995) Euphytica. 85:29-34; EP-A 0 223 247). Most preferred are biolistic transformation methods. Here, the DNA to be transformed 15 is applied to, for example gold or tungsten particles. These particles are subsequently accelerated towards the explant to be transformed (Dix PJ and Kavanagh TA (1995) Euphytica. 85:29-34; EP-A 0 223 247). Thereafter, transplastomic plants are regenerated under selection pressure on suitable medium in the 20 manner with which the skilled worker is familiar. Such methods have been described (for example US 5,451,513; US 5,877,402; Svab Z et al. (1990) Proc Natl Acad Sci USA 87:8526-8530; Svab Z and Maliga P (1993) Proc Natl Acad Sci USA 90:913-917). Moreover, the DNA can be introduced into the plastids by means of 25 microinjection. A specific microinjection method has been described recently (Knoblauch M et al. (1999) Nature Biotech 17:906-909; van Bel AJE et al. (2001) Curr Opin Biotechnol 12:144-149). This method is particularly preferred for the present invention. It is also possible to introduce, by means of 30 protoplast fusion, the plastids from one species into a different species, to transform them in the latter and subsequently to return them to the original species by protoplast fusion (WO 01/70939).

35 Besides these "direct" transformation techniques, transformation can also be carried out by bacterial infection by means of *Agrobacterium tumefaciens* or *Agrobacterium rhizogenes* (Horsch RB (1986) Proc Natl Acad Sci USA 83(8):2571-2575; Fraley et al. (1983) Proc Natl Acad Sci USA 80:4803-4807; Bevens et al. (1983) 40 Nature 304:184-187). The expression cassette for, for example, the DSB1 enzyme is preferably integrated into specific plasmids, either into a shuttle, or intermediate, vector or into a binary vector. Binary vectors are preferably used. Binary vectors are capable of replication both in *E. coli* and in *Agrobacterium* and 45 be transformed directly into *Agrobacterium* (Holsters et al. (1978) Mol Gen Genet 163:181-187). Various binary vectors are known; some of them are commercially available such as, for

example, pBIN19 (Clontech Laboratories, Inc. USA; Bevan et al. (1984) Nucl Acids Res 12:8711). The selection marker gene permits the selection of transformed agrobacteria and is, for example, the nptII gene, which confers resistance to kanamycin. The binary  
5 plasmid can be transferred into the agrobacterial strain for example by electroporation or other transformation methods (Mozo & Hooykaas 1991, Plant Mol. Biol. 16, 917-918). The plant explants are generally cocultured with the agrobacterial strain for two to three days. The agrobacterium which, in this case,  
10 acts as the host organism, should already comprise a plasmid with the vir region. Many Agrobacterium tumefaciens strains are capable of transferring genetic material, such as, for example, the strains EHA101[pEHA101] (Hood EE et al. (1996) J Bacteriol 168(3):1291-1301), EHA105[pEHA105] (Hood et al. (1993) Transgenic  
15 Research 2:208-218), LBA4404[pAL4404] (Hoekema et al. (1983) Nature 303:179-181), C58C1[pMP90] (Koncz and Schell (1986) Mol Gen Genet 204:383-396) and C58C1[pGV2260] (Deblaere et al. (1985) Nucl Acids Res 13: 4777-4788).

20 To transfer the DNA to the plant cell, plant explants are cocultured with Agrobacterium tumefaciens or Agrobacterium rhizogenes. Starting from infected plant material (for example leaf, root or stem portions, but also protoplasts or suspensions of plant cells), intact plants can be regenerated using a  
25 suitable medium which may comprise, for example, antibiotics or biocides for selecting transformed cells. A cotransformed selection marker permits the selection of transformed cells from untransformed cells (McCormick et al. (1986) Plant Cell Reports 5:81-84). The plants obtained can be bred, selfed and hybridized  
30 in the customary manner. Two or more generations should be grown to ensure that the genomic integration is stable and hereditary. The abovementioned methods are described in, for example, Jenes B et al. (1993) Techniques for Gene Transfer, in: Transgenic Plants, Vol. 1, Engineering and Utilization, edited by Kung SD  
35 and Wu R, Academic Press, pp.128-143 and in Potrykus (1991) Ann Rev Plant Physiol Plant Molec Biol 42:205-225).

The Agrobacterium-mediated transformation is best suited to dicotyledonous plant cells, whereas the direct transformation  
40 techniques are suitable for any cell type.

The Agrobacterium-mediated transformation is especially preferably employed for the transformation of the nucleus, while the direct transformation techniques are especially preferably  
45 employed for the transformation of plastids.

## 70

As soon as a predominantly homotransplastomic plant cell has been generated by the method according to the invention, an intact plant can be obtained using methods with which the skilled worker is familiar. The starting material for this purpose is, for  
5 example, callus cultures. The development of shoot and root can be induced in the known manner in this as yet undifferentiated biomass. The shoots obtained can be planted out and used for breeding.

## 10 Deletion methods

In the above-described methods according to the invention, it is advantageous, at various levels, to remove certain sequences which have previously been introduced (for example sequences for  
15 selection markers and/or DSB1 enzymes) from the plastome or genome of the plant or cell. Thus, it is advantageous, but not necessarily required, to remove the selection marker which has been introduced for example during the insertion of a nonnatural DSB recognition sequence, from the master plant since the same  
20 selection marker can then be utilized in a subsequent transformation (for example with the insertion sequence). Deletion is particularly advantageous since the selection marker is no longer necessarily required after the selection phase, and therefore superfluous. Moreover, deletion increases the consumer  
25 acceptance and is desirable for registration purposes. Moreover, the protein synthesis apparatus of the plastid is not unnecessarily burdened by the synthesis of the marker protein, which has potentially advantageous effects on the characteristics of the plant in question.

30

The skilled worker is familiar with a variety of methods for the directed deletion of sequences. One which should be mentioned by way of example, but not by limitation, is the excision by means of recombinases. Various sequence-specific recombination systems  
35 have been described, such as the Cre/lox system of bacteriophage P1 (Dale EC and Ow DW (1991) Proc Natl Acad Sci USA 88:10558-10562; Russell SH et al. (1992) Mol Gen Genet 234: 49-59; Osborne BI et al. (1995) Plant J. 7, 687-701), the yeast FLP/FRT system (Kilby NJ et al. (1995) Plant J 8:637-652; Lyznik  
40 LA et al. (1996) Nucleic Acids Res 24:3784-3789), the Gin recombinase of the phage Mu, the Pin recombinase from E. coli or the R/RS system of the plasmid pSR1 (Onouchi H et al. (1995) Mol Gen Genet 247:653-660; Sugita K et al. (2000) Plant J 22:461-469). These methods can be utilized not only for deleting  
45 DNA sequences from the nuclear genome, but also from the plastome (Corneille et al. (2001) Plant J 27: 171-178; Hajdukiewicz et al. (2001) Plant J 27:161-170). Further recombinases which can be

employed are, for example, PhiC31 (Kuhstoss & Rao (1991) J Mol Biol 222:897-908), TP901 (Christiansen et al. (1996) J Bacteriol 178:5164-5173), xisF from Anabaena (Ramaswamy et al. (1997) Mol Microbiol 23:1241-1249), integrase from phage PhiLC3 (Lillehaug et al. (1997) Gene 188:129-136) or the recombinase encoded by the sre gene of the R4 phage (Matsuura et al. (1996) J Bacteriol 178:3374-3376).

In a preferred embodiment, however, the deletion is effected by intrachromosomal recombination owing to suitably introduced sequence duplications. The efficiency of the latter can be enhanced by the directed introduction of double-strand breaks near the sequence duplications (cf. Fig. 8). To this end, the sequence to be deleted is flanked bilaterally by homology sequences H1 and H2 which have sufficient length and homology to undergo recombination with one another. Recombination is induced by the induction of at least one sequence-specific double-strand break of a further DSB recognition sequence located near one of the two homology sequences (but preferably different from the first one). This DSB recognition sequence is preferably localized between the two homology sequences. To induce the double-strand break, it is preferred to express or introduce a second DSBI enzyme which differs from the first one. This method is especially preferably utilized for deleting selection markers from the plastome.

The invention furthermore relates to the transplastomic, predominantly homoplastomic, plants generated using the method according to the invention, and to parts of these plants, such as leaves, roots, seeds, fruits, tubers, pollen or cell cultures, callus and the like which are derived from such plants.

The invention furthermore relates to the plants employed in the method according to the invention which comprise an expression cassette according to the invention for a DSBI enzyme or a fusion protein of PLS and DSBI enzyme. In this context, the expression cassette for the fusion protein of PLS and DSBI enzyme is - especially preferably - stably integrated in the nuclear DNA under the control of a promoter which is functional in the plant nucleus. The expression cassette encoding a DSBI enzyme under the control of a promoter which is active in plant plastids is preferably stably integrated into the plastome. Comprised are furthermore parts of same such as leaves, roots, seeds, tubers, fruits, pollen or cell cultures, callus and the like which are derived from the abovementioned plants.

## 72

Genetically modified plants according to the invention which can be consumed by humans and animals can also be used as foods or feeds, for example directly or following processing known per se.

- 5 The invention furthermore relates to the use of the above-described transplastomic, predominantly homoplastomic, plants according to the invention and of the cells, cell cultures, parts – such as, for example, the roots, leaves and the like in the case of transgenic plant organisms – and transgenic  
10 propagation material such as seeds or fruits which are derived from them for the production of foods or feeds, pharmaceuticals or fine chemicals.

- Fine chemicals refers to enzymes such as, for example, the  
15 industrial enzymes mentioned hereinbelow, vitamins such as, for example, tocopherols and tocotrienols (for example vitamin E) and vitamin B2, amino acids such as, for example, methionine, lysine or glutamate, carbohydrates such as, for example, starch, amylose, amylopectin or sucrose, fatty acids such as, for  
20 example, saturated, unsaturated and polyunsaturated fatty acids, natural and synthetic flavorings, aroma chemicals such as, for example linalool, menthol, borneone (camphor), pinene, limonene or geraniol, and colorants such as, for example, retinoids (for example vitamin A), flavonoids (for example quercetin, rutin,  
25 tangeretin, nobiletin) or carotenoids (for example  $\beta$ -carotene, lycopene, astaxanthin). The production of tocopherols and tocotrienols and of carotenoids is especially preferred. Growing the transformed host organisms, and isolation from the host organisms or the growth medium, are carried out using methods  
30 with which the skilled worker is familiar. The production of pharmaceuticals such as, for example, antibodies or vaccines has been described (Hood EE, Jilka JM. (1999) Curr Opin Biotechnol. 10(4):382-386; Ma JK and Vine ND (1999) Curr Top Microbiol Immunol. 236:275-92).

- 35 The method according to the invention is particularly suitable for producing industrial enzymes within what is known as "phytofarming". Examples of industrial enzymes which may be mentioned, but not by way of limitation, are lipases, esterases,  
40 proteases, nitrilases, acylases, epoxyhydrolases, amidases, phosphatases, xylanases, alcohol dehydrogenases, amylases, glucosidases, galactosidases, pullulanases, endocellulases, glucanases, cellulases, nucleases, chitin deacetylases, monoaminoxidases, lysozymes and laccases.

## 73

Embodiments which are especially preferred for the purposes of the invention are described in greater detail hereinbelow within the explanations of the figures.

## 5 Sequences

1. SEQ ID NO: 1  
PCB42-94 Basic vector for plastid transformation.
- 10 2. SEQ ID NO:2  
Nucleic acid sequence inserted into the multiple cloning site of PCB42-94 (SEQ ID NO: 1). Resulting vector: PCB199-3.
- 15 3. SEQ ID NO:3  
Nucleic acid sequence inserted into the multiple cloning site of PCB42-94 (SEQ ID NO: 1). Resulting vector: PCB401-20
- 20 4. SEQ ID NO:4  
Expression cassette from PCB289-13 for the expression of the I-PpoI homing endonuclease in plastids.
- 25 5. SEQ ID NO:5  
Amino acid sequence of the I-PpoI homing endonuclease encoded by the expression cassette from PCB289-13.
6. SEQ ID NO:6  
XhoI / BglII fragment employed for generating the vector PCB304-25.
- 30 7. SEQ ID NO:7  
Nucleic acid sequence inserted into the multiple cloning site of pGEMTeasy. Resulting vector: PCB220-17
- 35 8. SEQ ID NO:8  
Nucleic acid sequence inserted into the multiple cloning site of pBluescript. Resulting vector: PCB270-1
- 40 9. SEQ ID NO:9  
Sequence from vector PCB315-1: LacZ gene with inserted intron for detecting splicing.
10. SEQ ID NO: 10  
L1.LtrB intron from vector PCB345-34.

## 74

11. SEQ ID NO: 11  
Synthetic sequence of the homing endonuclease I-PpoI (ORF: 16 to 507)
- 5 12. SEQ ID NO: 12  
Protein sequence of the homing endonuclease I-PpoI
13. SEQ ID NO: 13  
Nucleic acid sequence of the homing endonuclease I-CpaI from  
10 *Chlamydomonas pallidostigmatica* (modification of the published sequence at position 69. An NcoI cleavage site was introduced at ATG (ORF: 4 to 462)
14. SEQ ID NO: 14  
15 Protein sequence of the homing endonuclease I-CpaI
15. SEQ ID NO: 15  
Sequence comprising the CpLSU2 intron
- 20 16. SEQ ID NO: 16: Oligonucleotide primer p19  
5'-TAAGGCCCTCGGTAGCAACGG-3'
17. SEQ ID NO: 17: Oligonucleotide primer p20  
5'-GGGGTACCAAATCCAACACTAG-3'
- 25 18. SEQ ID NO: 18: Oligonucleotide primer p21:  
5'-GGAGCTCGCTCCCCGCCGTCGTTC-3'
19. SEQ ID NO: 19: Oligonucleotide primer p22  
30 5'-GATGCATGATGACTTGACGGCATCCTC-3'
20. SEQ ID NO: 20: Oligonucleotide primer p190  
5'-GTCGACAGATCTTTAA-3'
- 35 21. SEQ ID NO: 21: Oligonucleotide primer p191  
5'-AGATCTGTCTGACTTAA-3'
22. SEQ ID NO: 22: Oligonucleotide primer p199  
5'-GATCTCCAGTTAACTGGGGTAC-3'
- 40 23. SEQ ID NO: 23: Oligonucleotide primer p200  
5'-CCCAGTTAACTGGA-3'
24. SEQ ID NO: 24: Oligonucleotide primer p218  
45 5'-TTAAGCCAGTTAACTGGGCGGAGCT-3'

## 75

25. SEQ ID NO: 25: Oligonucleotide primer p219  
5'-CCGCCCAGTTAACTGGC-3'
26. SEQ ID NO: 26: Oligonucleotide primer p276  
5'-TCGAGAAGATCAGCCTGTTATCCCTAGAGTAACT-3'
27. SEQ ID NO: 27: Oligonucleotide primer p277  
5'-CTAGAGTTACTCTAGGGATAACAGGCTGATCTTC-3'
28. SEQ ID NO: 28: Oligonucleotide primer p91  
5'-AGAAGACGATCCTAAGG-3'
29. SEQ ID NO: 29: Oligonucleotide primer p92  
5'-TGAAGACTTGACAAGGAATTTTCGC-3'
30. SEQ ID NO: 30: Oligonucleotide primer p102  
5'-AGAAGACGATCCTAAATAGCAATATTTACCTTTGGGACCAAAAGTTATCAGGCATG-3
31. SEQ ID NO: 31: Oligonucleotide primer p103  
5'TGAAGACTTGACAAGGAATTTGCTACCTTCGAGTACTCCAAACTAATC-3'
32. SEQ ID NO: 32: Oligonucleotide primer p207  
5'-GAGAAGACATTCTAACACATCCATAACGTGCG-3'
33. SEQ ID NO: 33: Oligonucleotide primer p208  
5'-TGAAGACTTGACATTTGATATGGTGAAGTAGG-3'
34. SEQ ID NO: 34  
Nucleic acid sequence encoding the transit peptide of the  
small subunit (SSU) of ribulose biphosphate carboxylase  
(Rubisco ssu) from pea
35. SEQ ID NO: 35  
Transit peptide of the small subunit (SSU) of ribulose  
biphosphate carboxylase (Rubisco ssu) from pea
36. SEQ ID NO: 36  
Transit peptide of the tobacco plastidic transketolase.
37. SEQ ID NO: 37  
Nucleic acid sequence encoding the transit peptide of the  
tobacco plastidic transketolase (reading frame 1; pTP09)
38. SEQ ID NO: 38  
Nucleic acid sequence encoding the transit peptide of the  
tobacco plastidic transketolase (reading frame 2; pTP10)



39. SEQ ID NO: 39  
Nucleic acid sequence encoding the transit peptide of the tobacco plastidic transketolase (reading frame 3; pTP11)
- 5 40. SEQ ID NO: 40  
Transit peptide of the plastidic isopentenyl-pyrophosphate isomerase-2 (IPP-2) from *Arabidopsis thaliana*.
41. SEQ ID NO: 41  
10 Nucleic acid sequence encoding the transit peptide of the plastidic isopentenyl-pyrophosphate isomerase-2 (IPP-2) from *Arabidopsis thaliana* (reading frame 1; IPP-9)
42. SEQ ID NO: 42  
15 Nucleic acid sequence encoding the transit peptide of the plastidic isopentenyl-pyrophosphate isomerase-2 (IPP-2) *Arabidopsis thaliana* (reading frame 2; IPP-10)
43. SEQ ID NO: 43  
20 Nucleic acid sequence encoding the transit peptide of the plastidic isopentenyl-pyrophosphate isomerase-2 (IPP-2) from *Arabidopsis thaliana* (reading frame 3; IPP-11)
44. SEQ ID NO: 44  
25 Nucleic acid sequence encoding the tobacco PrbcL promoter.
45. SEQ ID NO: 45  
Nucleic acid sequence encoding the tobacco Prps16-107 promoter.  
30
46. SEQ ID NO: 46  
Nucleic acid sequence encoding the tobacco Prn16 promoter.
47. SEQ ID NO: 47  
35 Nucleic acid sequence encoding the tobacco PaccD-129 promoter.
48. SEQ ID NO: 48  
Nucleic acid sequence encoding the tobacco PclpP-53 promoter.  
40
49. SEQ ID NO: 49  
Nucleic acid sequence encoding the tobacco Prn-62 promoter.
50. SEQ ID NO: 50  
45 Nucleic acid sequence encoding the tobacco Prps16 promoter.

## 77

51. SEQ ID NO: 51  
Nucleic acid sequence encoding the tobacco PatpB/E-290 promoter.
- 5 52. SEQ ID NO: 52  
Nucleic acid sequence encoding the tobacco PrpoB-345 promoter.
53. SEQ ID NO: 53  
10 Nucleic acid sequence encoding a promoter derived from the consensus sequence of the E. coli  $\sigma$ 70 promoters.
54. SEQ ID NO: 54  
Nucleic acid sequence encoding the 5'-untranslated region of  
15 the tobacco psbA gene (5'psbA)
55. SEQ ID NO: 55  
Nucleic acid sequence encoding the 5'-untranslated region  
including 5' portions from the coding region of the tobacco  
20 rbcL gene (5'rbcL).
56. SEQ ID NO: 56  
Nucleic acid sequence encoding the 5'-untranslated region of  
the tobacco rbcLs gene.  
25
57. SEQ ID NO: 57  
Nucleic acid sequence encoding the 3'-untranslated region of  
the Synechocystis psbA-1 gene (3'psbA-1)
- 30 58. SEQ ID NO: 58  
Nucleic acid sequence encoding the 3'-untranslated region of  
the tobacco psbA gene (3'psbA)
59. SEQ ID NO: 59  
35 Nucleic acid sequence encoding the 3'-untranslated region of  
the tobacco rbcL gene (3'rbcL)
60. SEQ ID NO: 60  
Nucleic acid sequence encoding synthetic ribosome binding  
40 sites (RBS)
61. SEQ ID NO: 61  
Nucleic acid sequence encoding synthetic ribosome binding  
sites (RBS)  
45

## 78

62. SEQ ID NO: 62  
Complete insert of the vector pCB304-25
63. SEQ ID NO: 63  
5 BglIII / MunI fragment of the vector pCB320-192.
64. SEQ ID NO: 64: Oligonucleotide primer p93  
5'-AAAGATCTCCTCACAAAGGGGGTCG-3'
- 10 65. SEQ ID NO: 65: Oligonucleotide primer p97  
5'-TCGAAGACTTAGGACCGTTATAG-3'
66. SEQ ID NO: 66: Oligonucleotide primer p98  
5'-AGGAAGACCTTGTCGGGTAAGTTCCG-3'
- 15 67. SEQ ID NO: 67: Oligonucleotide primer p95:  
5'-CTCAATTGGGGTCTCTGTCCAGGTGCAGG-3'
68. SEQ ID NO: 68: Nucleic acid sequence encoding fusion proteins  
20 from the native I-Ppo-I nuclease and the IPP plastid  
localization sequence (ORF for I-PpoI: 181-672; IPP transit  
peptide: 1-180; native sequence from 1-172).
69. SEQ ID NO: 69: Fusion proteins of the native I-Ppo-I nuclease  
25 and the IPP-plastid localization sequence.
70. SEQ ID NO: 70: Nucleic acid sequence encoding long version of  
the I-PpoI homing endonuclease.
- 30 71. SEQ ID NO: 71: Amino acid sequence encoding long version of  
the I-PpoI homing endonuclease.
72. SEQ ID NO: 72: Nucleic acid sequence encoding a  
promoter sequence derived from the  
35 consensus sequence of the  $\sigma$ 70 promoters  
from E.coli.
73. SEQ ID NO: 73: Nucleic acid sequence encoding the  
artificial intron TetIVS2a.
- 40 74. SEQ ID NO: 74: Insert of vector pCB459-1
75. SEQ ID NO: 75: Insert of vector pCB478-3
- 45 76. SEQ ID NO: 76: Insert of vector pCB492-25

## 79

77. SEQ ID NO: 77: Oligonucleotide primer p396  
 5'-TAGTAAATGACAATTTTCCTCTGAATTATATAATTAACATGGCGACTGTTTACCAAAA  
 AC-3

5 78. SEQ ID NO: 78: Oligonucleotide primer p95  
 5'-CTCAATTGGGGTCTCTGTCCAGGTGCAGG-3'

79. SEQ ID NO: 79: Nucleic acid sequence encoding PCR product  
 Prom-TetIVS2a-Cpa

10

80. SEQ ID NO: 80: Insert of vector pCB435-45

81. SEQ ID NO: 81: Nucleic acid sequence encoding probe for  
 Southern blot analysis (directed against  
 15 portions of the 16SrDNA).

82. SEQ ID NO: 82: Nucleic acid sequence encoding probe for  
 Southern blot analysis (directed against  
 20 portions of the 23SrDNA).

83. SEQ ID NO: 83: Insert of vector pCB456-2

84. SEQ ID NO: 84: Insert of vector pCB528-2 from KpnI to SacI

## 25 Figures

Within the method according to the invention, particularly the  
 embodiments detailed in the figures hereinbelow are especially  
 preferred. The following abbreviations are generally used in the  
 30 figures:

- A, A' Pair of homologous sequences A and A'  
 A/A' Result of a homologous recombination between A and A'  
 and/or a substitution of A by A' caused by repair  
 35 synthesis.
- B, B' Pair of homologous sequences B and B'  
 B/B' Result of a homologous recombination between B and B'  
 and/or a substitution of B by B' caused by repair  
 synthesis.
- 40 H1, H2: Pair of homologous sequences H1 and H2  
 H1/2: Sequence as the result of the homologous recombination of  
 H1 and H2
- DS Functional DSB recognition sequence  
 nDS nonfunctional half of a DSB recognition sequence
- 45 E: DSBI enzyme  
 P: Promoter  
 I: Further nucleic acid sequence (gene of interest)

## 80

S, S' Positive selection markers  
NS Negative selection marker  
IS Intron sequences. The intron in total is marked as a box.  
The box comprises all elements required for a functional  
5 intron.

As already described above, A/A' and B/B' are the result of a homologous recombination and/or a substitution brought about by repair synthesis. The resulting sequence, in turn, can be the  
10 starting sequence for further homologous recombinations or repair syntheses. For the sake of simplicity, this sequence (A/A' and B/B') is again referred to as A and B, respectively, in the steps which follow.

15 1. Fig. 1: Introduction of a DSB recognition sequence into the plastome by means of double cross-over

In an especially preferred embodiment 1, a DSB construct is first introduced into plastids of a higher plant. In this  
20 embodiment, the DSB construct is preferably equipped with homologous target regions and with an expressible selection marker (promoter - 5'UTR - selection marker - 3'UTR) and, in this embodiment, preferably comprises a recognition region for a DSB enzyme which preferably has no natural recognition sequence in  
25 the plastidic genome of the (untransformed) plant in question. The DSB construct can optionally already encode further genes of interest. Predominantly homoplastomic master plants are generated (Fig. 1).

30 2. Fig. 2A-E: Introduction of an insert sequence with an expression cassette for a DSB enzyme and, if appropriate, selection markers and further genes of interest

Explants of the master plants generated in embodiment 1 are  
35 utilized for a further transformation with a transformation construct according to the invention. Preferably, the transformation construct according to the invention has regions which are homologous to the sequences surrounding the insertion site of the DSB construct, which regions are preferably located  
40 on both sides (Fig. 2A, 2B) or on one side (Fig. 2C, 2D) of the insertion sequence. Insertion now takes place via homologous recombination (for example cross-over) or via repair synthesis.

The sequences to be inserted are especially preferably -  
45 following the homology sequences inwardly - flanked by portions of the DSB recognition sequence (nDS) which correspond to the portions originating as the consequence of cleavage with the DSB

## 81

enzyme (Fig. 2B). The insertion sequence thus comprises sequences which correspond in detail to the ends which are the result of a cleavage in the plastome and thus ensure a particularly efficient incorporation.

5

If the transformation construct, or the insertion sequence, has no such homologous regions, the insertion sequence is preferably provided, at these ends, with overhangs which are also generated by the DSBI enzyme after cleavage of the master plant plastome 10 (Fig. 2E).

If only one homology sequence is present, this sequence borders, in an especially preferred embodiment, an nDS sequence (see hereinabove as described for Fig. 2B), while the other side of 15 the insertion sequence is provided with overhangs which correspond to those generated by the DSBI enzyme in the plastome of the master plant (Fig. 2D).

The insertion sequence optionally codes for a further expressible 20 selection marker which differs functionally from that of the DSBR construct, if appropriate one or more expressible genes of interest and the expressible DSBI enzyme, which cleaves the recognition sequence introduced by the DSBR construct at the insert site in the plastidic genome of the master plant. The 25 insertion sequence of the transformation construct is inserted, in this context, in such a way at a position that said recognition region is no longer functional after the insertion.

3. Fig.3: Introduction of a DSB recognition sequence into the 30 plastome by means of double cross-over in a transcriptionally active region

In a further, especially preferred embodiment 2, a DSBR construct is initially introduced into plastids of a higher plant. In this 35 embodiment, the DSBR construct is preferably equipped with homologous target regions and with an expressible selection marker, an endogenous promoter of the plastome being utilized, and it preferably additionally comprises a recognition sequence for a DSBI enzyme which preferably has no natural recognition 40 sequence in the plastidic genome of the (nontransformed) plant in question. The DSBR construct may already encode genes of interest. Predominantly homoplastomic master plants are generated (Fig. 3).

45 4. Fig.4A-E: Introduction of an insertion sequence with a cassette encoding a DSBI enzyme and, if appropriate, a

## selection marker and further genes of interest

Explants of the master plants generated in embodiment 2 are utilized for a further transformation with a transformation  
5 construct according to the invention. Preferably, the transformation construct according to the invention has regions which are homologous to the sequences surrounding the insertion site of the DSBR construct, which regions are preferably located on both sides (Fig. 4A, 4B) or on one side (Fig. 4C, 4D) of the  
10 insertion sequence. Insertion now takes place via homologous recombination (for example cross-over) or via repair synthesis. The sequences to be inserted are especially preferably – following the homology sequences inwardly – flanked by portions of the DSBR recognition sequence (nDS) which correspond to the  
15 portions originating as the consequence of cleavage with the DSBI enzyme (Fig. 4B). The insertion sequence thus comprises sequences which correspond in detail to the ends which are the result of a cleavage in the plastome and thus ensure a particularly efficient incorporation.

20 If the transformation construct, or the insertion sequence, has no such homologous regions, the insertion sequence is preferably provided, at these ends, with overhangs which are also generated by the DSBI enzyme after cleavage of the master plant plastome  
25 (Fig. 4E).

If only one homology sequence is present, this sequence borders, in an especially preferred embodiment, an nDS sequence (see hereinabove as described for Fig. 4B), while the other side of  
30 the insertion sequence is provided with overhangs which correspond to those generated by the DSBI enzyme in the plastome of the master plant (Fig. 4D).

The insertion sequence optionally codes for a further expressible  
35 selection marker which differs functionally from that of the DSBR construct, if appropriate one or more expressible genes of interest and the expressible DSBI enzyme, which cleaves the recognition sequence introduced by the DSBR construct at the insert site in the plastidic genome of the master plant. The  
40 insertion sequence of the transformation construct is inserted, in this context, in such a way at a position that said recognition region is no longer functional after the insertion.

45 5. Fig.5A-E: Introduction of an insertion sequence with a cassette encoding a DSBI enzyme and, if appropriate, selection markers and further genes of interest utilizing

## 83

## natural, endogenous DSB recognition sequences

In a further, very especially preferred embodiment 3, a transformation construct according to the invention comprises an expressible DSBI enzyme which has an endogenous, natural recognition sequence in the plastome of the plant in question.

Explants of these natural master plants are utilized for a transformation with a transformation construct according to the invention. Preferably, the transformation construct according to the invention has regions which are homologous to the sequences surrounding the insertion site of the DSB construct, which regions are preferably located on both sides (Fig. 5A, 5B) or on one side (Fig. 5C, 5D) of the insertion sequence. Insertion now takes place via homologous recombination (for example cross-over) or via repair synthesis.

The sequences to be inserted are especially preferably - following the homology sequences inwardly - flanked by portions of the DSB recognition sequence (nDS) which correspond to the portions originating as the consequence of cleavage with the DSBI enzyme (Fig. 5B). The insertion sequence thus comprises sequences which correspond in detail to the ends which are the result of a cleavage in the plastome and thus ensure a particularly efficient incorporation.

If the transformation construct, or the insertion sequence, has no such homologous regions, the insertion sequence is preferably provided, at these ends, with overhangs which are also generated by the DSBI enzyme after cleavage of the master plant plastome (Fig. 5E).

If only one homology sequence is present, this sequence borders, in an especially preferred embodiment, an nDS sequence (see hereinabove as described for Fig. 5B), while the other side of the insertion sequence is provided with overhangs which correspond to those generated by the DSBI enzyme in the plastome of the master plant (Fig. 5D).

The insertion sequence of the transformation construct is, in this context, preferably inserted at a position in such a way that said recognition region is no longer functional after the insertion. The insertion sequence preferably encodes an expressible selection marker (S'), one or more genes of interest, and the expressible DSBI enzyme. The selection marker is optional.



## 84

6. Fig.6A-E: Introduction of an insertion sequence with a cassette encoding genes of interest and, if appropriate, selection markers, and introduction of a DSBI enzyme in trans

5

In further preferred embodiments 4, the DSBI enzyme is not encoded by the transformation construct, but either expressed in trans (in plastids or as PLS fusion protein in the nucleus) or transfected into the plastids in the form of RNA or protein. The DSBI enzyme recognizes either an artificially introduced DSB recognition sequence (Fig. 6A, 6B) or a natural DSB recognition sequence (Fig. 6C, 6D). This embodiment is especially preferred when the transformation construct comprises no promoter elements, and expression of the coded genes is only realized after insertion into the plastome, using plastidic, endogenous promoters.

As was the case in the embodiments which have already been described above, the transformation construct preferably has regions bilaterally (Fig. 6A, 6B) or unilaterally (not shown) of the insertion sequence which are homologous to the sequences surrounding the insertion site of the DSBR construct. Insertion now takes place via homologous recombination (for example cross-over) or repair synthesis.

25

The sequences to be inserted are especially preferably – following the homology sequences inwardly – flanked by portions of the DSB recognition sequence (nDS) which correspond to the portions originating as the consequence of cleavage with the DSBI enzyme (Fig. 6B, 6D). The insertion sequence thus comprises sequences which correspond in detail to the ends which are the result of a cleavage in the plastome and thus ensure a particularly efficient incorporation.

35 If the transformation construct, or the insertion sequence, has no such homologous regions, the insertion sequence is preferably provided, at these ends, with overhangs which are also generated by the DSBI enzyme after cleavage of the master plant plastome (Fig. 6E). The transformation construct can additionally comprise a sequence encoding a DSBI enzyme. However, expression only takes place after successful insertion into the plastome, so that it is desirable that a first amount of functional RNA or protein of a DSBI enzyme is provided.

45 If only one homology sequence is present, it borders, in an especially preferred embodiment, an nDS sequence (see above as described for Fig. 6B, 6D), while the other side of the insertion

## 85

sequence is provided with overhangs which correspond to those generated by the DSBI enzyme in the plastome of the master plant (not shown).

- 5 In this context, the insertion sequence of the transformation construct is preferably inserted at a position in such a way that said recognition region is no longer functional after the insertion.
- 10 7. Fig.7A-E: Introduction of an insertion sequence comprising an intron sequence with a cassette encoding genes of interest and, if appropriate, selection markers or DSBI enzymes
- 15 In a further, very especially preferred embodiment 5, the gene of interest (and optionally a selection marker S' and/or the DSBI enzyme) is/are encoded within an intron which is functional at the insertion site selected, i.e. which can splice out of the transcript formed therein.
- 20 Preferably, the transformation construct according to the invention has regions which are homologous to the sequences surrounding the insertion site of the DSBR construct, which regions are preferably located on both sides (Fig. 7A, 7B) or one
- 25 one side (not shown) of the insertion sequence. Insertion now takes place via homologous recombination (for example cross-over) or via repair synthesis.
- If the transformation construct or the insertion sequence has no
- 30 such homologous regions, the insertion sequence is preferably provided, at these ends, with overhangs which are also generated by the DSBI enzyme after cleaving the plastome of the master plant (not shown).
- 35 Expression can be controlled by means of a promoter which is present on the transformation construct (Fig. 7B) or an endogenous, plastidic promoter (Fig. 7A). In the first case, the DSBI enzyme is preferably present on the transformation construct (Fig. 7B), while, in the latter case, it is either expressed (at
- 40 least in parallel) in trans (in plastids or as PLS fusion protein in the nucleus) or transfected into the plastids in the form of RNA or protein (Fig. 7A).
- In this context, the insertion sequence of the transformation
- 45 construct is preferably inserted at a position in such a way that said recognition region is no longer functional after the insertion. The transformation construct can additionally

## 86

optionally comprise a sequence encoding a DSB1 enzyme. However, expression only takes place after successful insertion into the plastome, so that it is desirable that a first amount of functional RNA or protein of a DSB1 enzyme is provided.

5

8. Fig.8: Deletion of sequences by means of intramolecular homologous recombination induced by sequence-specific double-strand breaks

10 In all of the above-described embodiments, sequences – for example those encoding selection markers or DSB1 enzymes – are preferably flanked by homology sequences H1 and H2 with sufficient length and homology to undergo recombination with one another. The recombination is induced by the induction of at  
15 least one double-strand break in the DSB recognition sequence located between the two homology sequences. To induce the double-strand break, it is preferred to transiently express or introduce a DSB1 enzyme (Fig. 8).

20 The skilled worker realizes that the sequence of the genes expressed in an operon is exchangeable and can thus vary in the above-described embodiments. Also, when using only one homology sequence for inserting the insertion sequence, this homology sequence may be localized at the 5' side (as shown in the  
25 figures) or the 3' side of the double-strand break. In principle, the DSB1 enzyme can be expressed on the transformation construct and/or separately (in the nucleus or the plastids) and introduced differently into plastids, for example by transfection with RNA or protein.

30

9. Fig.9: Southern analysis of predominantly homotransplastomic plants

Wild-type and predominantly homotransplastomic master plants were  
35 analyzed with regard to the modification (introduction of a DSB recognition sequence; cf. Example 4). Owing to the modification, a 1750 bp band was detected (lanes 2, 3 and 4 corresponding to lines CB199NTH-4, -6 and -8), while a 3100 bp band was detected in the unmodified wild-type plant (lane 1).

40

10. Fig.10: Modification of the IGS of the Tetrahymena LSU intron.

Capital letters indicate the sequence of the intron, while  
45 lower-case letters represent the sequence of the surrounding exons. The flanking exon sequences, the 5' and 3' portion of the intron or intron derivative, and the sequence comprising the IGS

are shown. Bars between the bases indicate possible base pairings which can be formed for initiating the splicing procedure.

5 A: The abovementioned sequence segments of the naturally occurring *Tetrahymena* LSU intron in its natural exon environment are shown.

10 B: The abovementioned sequence segments of the *Tetrahymena* LSU intron derivative generated within the scope of the present invention (TetIVS2a) in the above-defined exon environment, as is found in the CpLSU5 intron within the DSB recognition sequence of the DSB1 enzyme I-CpaI are shown. Letters in bold represent the mutations carried out in comparison with the natural sequence.

15

11. Fig.11:

20 A: Southern analysis with BamHI-cut total DNA from the tobacco lines CB255+435NTH-16b, -16c, -19 and -20. A region of the 16Sr DNA was used as probe. The bands representing plastome copies which correspond to the wild type (WT; approx. 3.2 kb band detected) and those which bear the transgene (TG; approx. 2.3 kb band detected) are identified by arrows.

25 B: Schematic representation of transplastomic tobacco plants which have originated by the insertion of the insertion sequence from pCB435-45; and the bands to be expected in a corresponding Southern analysis (cf. A).

30 C: Southern analysis with HindIII-cut total DNA from the tobacco lines CB255+435NTH-16b, -16c, -19 and -20. A region of the 23Sr DNA was used as probe. The bands representing plastome copies which correspond to the wild type (WT; approx. 1.1 kb band detected) and those which bear the transgene (TG; approx. 1.5 kb band detected) are identified by arrows.

35

40 D: Schematic representation of transplastomic tobacco plants which have originated by the insertion of the insertion sequence from pCB255-1; and the bands to be expected in a corresponding Southern analysis (cf. C).

12. Fig. 12:

45 A: Wild-type and predominantly homotransplastomic master plants were analyzed in a Southern analysis with regard to the modification (introduction of one of the I-PpoI DSB recognition sequence; cf. Example 14.2). Owing to the

## 88

modification, an approximately 1.7 kb band was detected in the DNA which had been treated here with EcoRI (TG; lanes 1 and 4 corresponding to lines CB456NTH-1 and -15), while an approximately 3.1 kb band was detected in the unmodified wild-type plant (WT; lane 6). (wt - unmodified wild-type plant; wild type - shows the expected fragment size in unmodified wild-type plants; transgenic - shows the expected fragment size in plants CB456NTH)

10 B: Schematic representation of the EcoRI fragment which was to be expected in A by hybridization with the probe in a modified plant CB456NTH. (trnV - gene encoding a tRNA-Val; rrn16 - gene encoding the 16SrRNA; aadA - gene encoding a selection marker; 3'psbA (Synec) - noncoding region upstream of the Synechocystis psbA-1 gene, here incorporated into the expression cassette for the selection marker aadA; Psynth. - synthetic promoter derived from the consensus sequence for E. coli o70 promoters; DSB-R: DSB recognition sequence).

## 20 Examples

## General methods:

Oligonucleotides can be synthesized chemically for example in the known manner, using the phosphoramidite method (Voet, Voet, 2nd edition, Wiley Press New York, pages 896-897). The cloning steps carried out within the scope of the present invention such as, for example, restriction cleavages, agarose gel electrophoresis, purification of DNA fragments, transfer of nucleic acids to nitrocellulose and nylon membranes, linkage of DNA fragments, transformation of E. coli cells, bacterial cultures, propagation of phages, and the sequence analysis of recombinant DNA, are carried out as described by Sambrook et al. (1989) Cold Spring Harbor Laboratory Press; ISBN 0-87969-309-6. Recombinant DNA molecules are sequenced with an ALF-Express laser fluorescence DNA sequencer (Pharmacia, Uppsala, Sweden) following the method of Sanger (Sanger et al. (1977) Proc Natl Acad Sci USA 74:5463-5467).

40 Example 1: Generating a basic vector for the transformation of plastids

Initially, the selected target regions were cloned from the plastome of tobacco cv. SR1 by means of PCR. The left-hand target region was amplified using the primers p19 and p20.

89

p19: 5'-TAAGGCCCTCGGTAGCAACGG-3' (SEQ ID NO: 16)

p20: 5'-GGGGTACCAAATCCAACTAG-3' (SEQ ID NO: 17)

5 The primers p21 and p22 were used for amplifying the right-hand target region, the last-mentioned primer additionally introducing a spectinomycin resistance into the amplified part of the 16S rDNA, in addition to the SR1 resistance (binding-type marker).

10 p21: 5'-GGAGCTCGCTCCCCCGCCGTCGTTC-3' (SEQ ID NO: 18)

p22: 5'-GATGCATGATGACTTGACGGCATCCTC-3' (SEQ ID NO: 19)

The two amplified regions were cloned into pBluescript and  
15 pZeroBlunt, respectively, and sequenced. The left-hand and right-hand target regions were subsequently cloned into the backbone of the pUC19 vector. The cleavage sites Eco0109I and PvuII of the vector were used for this purpose. A multiple cloning site from pBluescript (from KpnI to SacI) was cloned  
20 between the left-hand and right-hand target region. This multiple cloning site resides between the two plastid-encoded genes trnV and rrn16 in the basic vector for the transformation of plastids. This basic vector for the transformation of plastids was named pCB42-94 (SEQ ID NO: 1). The vector comprises the following  
25 sequence elements:

- a) Position complementary to bp 55-1405: right-hand target region with the partial gene of the 16S rRNA (complementary to bp 56 to 1322). The latter comprises mutations for the  
30 streptomycin resistance (SR1, position bp 346) and spectinomycin resistance (SPC1, position bp 68).
- b) Position complementary to bp 2374 to 1510: left-hand target region comprising ORF131 (bp 1729 to 2124) and trnV gene  
35 inter alia (complementary to bp 1613 to 1542).
- c) Position bp 1404 to 1511: multiple cloning site
- d) Position bp 2629 to 3417: ampicillin resistance in the vector  
40 backbone

Example 2: Generation of a vector (pCB199-3) for introducing a  
non-naturally occurring recognition region for the  
45 homing endonucleases I-PpoI into the plastome of tobacco

## 90

Various elements were cloned one after the other into the multiple cloning site of the basic vector pCB42-94 (SEQ ID NO: 1) for the transformation of plastids:

- 5 a) *frt* recognition region (mutated, contains no *Xba*I cleavage site; complementary to 1307-1354)
- b) expression cassette for expressing the marker gene *aadA*, consisting of:
  - 10 i) promoter of the gene for 16S rRNA (complementary to 1191-1281)
  - ii) 5'-untranslated region of the tobacco *rbcl* gene  
15 (complementary to 1167-1184) including mutated 5' portions of the *rbcl* gene (duplication of 6 AS of the *rbcl* gene, partly mutated, as the consequence of the cloning strategy so that a fusion encoding a total of 12 amino acids (complementary to 1131-1166) with the  
20 subsequent element, the *aadA* gene, was formed)
  - iii) *aadA* gene (complementary to 336-1130)
  - iv) the 3' region of the *psbA* gene (complementary to 232-323)  
25
- c) core recognition region for the homing endonuclease I-PpoI (complementary to 176-190).

This vector, which is referred to as pCB199-3, comprises the  
30 abovementioned elements within the nucleic acid sequence with the SEQ ID NO: 2 instead of the multiple cloning site in the basic vector for the transformation of plastids. The sequence which replaces the complete MCS from *Kpn*I to *Sac*I is shown. However, there is no longer a *Kpn*I cleavage site in the sequence shown,  
35 owing the cloning strategy.

Example 3: Generation of a further vector (pCB401-20) for  
introducing a non-naturally occurring recognition  
region for the homing endonuclease I-PpoI into the  
40 plastome of tobacco

In contrast to the vector pCB199-3 described in Example 2, the vector described herein comprises no promoter and 3'UTR linked directly to the selection marker *aadA*. Rather, the expression of  
45 the *aadA* gene is controlled starting from the promoter of the *trnV* gene which is localized in the plastidic genome or in the left-hand target region upstream of the *aadA* gene. The purpose of

## 91

generating this vector was to avoid sequence duplication by exploiting regulatory regions from the tobacco plastidic genome. To this end, various elements were cloned one after the other into the multiple cloning site of the basic vector pCB42-94 for the transformation of plastids:

- a) ribosome binding site (complementary to bp 1033 to 1050)
- b) aadA gene (complementary to bp 238 to 1032)
- c) core recognition region for the homing endonuclease I-PpoI (complementary to bp 176 to 190)

The resulting vector also confers spectinomycin resistance in *E. coli*. This vector, which is referred to as pCB401-20, comprises the abovementioned elements within the nucleic acid sequence with the SEQ ID NO: 3 instead of the multiple cloning site in the basic vector pCB42-94 (SEQ ID NO: 1) for the transformation of plastids. Again, all of the sequence which replaces the MCS (from SacI to KpnI) is shown.

Example 4: Generation of predominantly homoplastomic tobacco master plants comprising a non-natural DSB recognition sequence

The plasmid pCB199-3 was introduced into the plastids of tobacco (*Nicotiana tabacum* cv. Petit Havana) as described hereinbelow. The regenerated plants were named CB199NTH. Independent lines were provided with different last numbers (for example CB199NTH-4).

The vector pCB401-20 is introduced analogously into the plastids of tobacco. Accordingly, the resulting plants are named CB401NTH.

First, leaf disks of diameter 2.0 to 2.5 cm were punched out of in-vitro-cultured plants, using a sterile cork borer, and placed upside-down on a Petri dish with bombardment medium [MS salts (Sigma-Aldrich): 4.3 g/l; sucrose: 30.0 g/l, Phyto-agar (Duchefa, P1003): 0.6% (w/v); pH 5.8; after autoclaving, 1.0 mg/l thiamine (Duchefa, T0614) and 0.1 g/l myo-inositol (Duchefa, I0609) were added]. The underside of the leaf, which faced away from the agar, was subsequently bombarded using the particle gun. To this end, the plasmid DNA to be transformed (isolated from *E. coli* using Nucleobond AX100 Macherey & Nagel) was first applied to gold particles 0.6  $\mu$ m in side by the following protocol ("coating"). First, 30 mg of powdered gold (BioRad) were taken up in ethanol. 60  $\mu$ l of the gold suspension were transferred into a



## 92

fresh Eppendorf tube and the gold particles were sedimented by centrifugation (for 10 seconds). The gold particles were washed twice in in each case 200 µl of sterile water and, after a further centrifugation step, taken up in 55 µl of water. The following 5 were added rapidly, with continuous mixing (vortexing):

- 5 µl plasmid DNA (1 µg/µl)
- 50 µl 2.5 M CaCl<sub>2</sub>
- 20 µl 0.1 M spermidine

10

The suspension was subsequently vortexed for a further 3 minutes and subsequently centrifuged briefly. The gold/DNA complexes which had sedimented were washed once or twice in in each case 200 µl of ethanol and, after a further centrifugation step, 15 finally taken up in 63 µl of ethanol. 3.5 µl (corresponding to 100 µg of gold) of this suspension were applied to a macrocarrier for each bombardment.

The particle gun (BioRad, PDS1000He) was prepared in accordance 20 with the manufacturer's instructions, and the leaf explants were bombarded with the gold/DNA complexes from a distance of 10 cm. The following parameters were used: vacuum: 27 inch Hg, pressure 1100 psi. After the bombardment, the explants were incubated for 2 days in controlled-environment cabinets (24°C, 16 h light, 8 h 25 darkness) and subsequently divided into segments approximately 0.5 cm<sup>2</sup> in size, using a surgical blade. These segments were then transferred to regeneration medium [bombardment medium supplemented with 1 mg/l 6-benzylaminopurine (BAP, Duchefa, B0904) and 0.1 mg/l naphthylacetic acid (NAA, Duchefa, N0903)] 30 supplemented with 500 mg/l spectinomycin (Duchefa, S0188) and incubated for 10 to 14 days under the abovementioned conditions in a controlled-environment cabinet. After this period of time had elapsed, the leaf segments were transferred to fresh regeneration medium supplemented with 500 mg/l spectinomycin. This procedure 35 was repeated until green shoots formed on the explants. The shoots were removed using a surgical blade and grown on growth medium (like bombardment medium, but with 10 g/l sucrose instead of 30 g/l sucrose) supplemented with 500 mg/l spectinomycin.

40 To obtain as predominantly homoplastomic plants as possible, it is optionally possible to excise explants from the regenerated plants themselves and to place them on regeneration medium with 1000 mg/l spectinomycin. Regenerating shoots are transferred into glass containers with growth medium supplemented with 500 to

45

## 93

1000 mg/l spectinomycin. After the plants have rooted, they are transferred into the greenhouse, where they are grown in soil until the seeds have matured.

- 5 When the transformation was carried out with the plasmid pCB199-3, 8 plants with resistance to spectinomycin were obtained. PCR and Southern analyses proved that three of these lines (lines CB199NTH-4, -6 and -8) have indeed incorporated the aadA gene into the plastidic genome.

10

- To perform a Southern analysis of the transplastomic plants, total DNA from leaves was isolated from transformed and untransformed plants with the aid of the GenElute Plant Genomic DNA Kit (Sigma). The DNA was taken up in 200  $\mu$ l of eluate. 86  $\mu$ l portions of this were treated with in each case 10  $\mu$ l of 10x restriction puffer and 40 U of restriction endonuclease and incubated for 4 to 8 hours at the temperature recommended for the restriction enzyme. The DNA was subsequently precipitated with ethanol in the manner known to the skilled worker and the precipitate was subsequently taken up in 20  $\mu$ l of water. The samples were subsequently separated on agarose gel by methods known to the skilled worker, and the DNA was denatured in the gel and transferred to a nylon membrane by means of a capillary blot.
- 25 A suitable probe for the radioactive hybridization was generated with the aid of the HighPrime (Roche) system. First, the membrane was prehybridized for 1 hour at 65°C with Hyb buffer (1% (w/v) bovine serum albumin; 7% (w/v) SDS; 1 mM EDTA; 0.5 M sodium phosphate buffer, pH 7.2). The heat-denatured probe was subsequently added and left to hybridize overnight at 65°C. The blots were subsequently washed as follows: one rinse with 2 x SSPE/0.1% SDS; washing for 15 minutes at 65°C with 2 x SSPE/0.1% SDS; washing for 15 minutes at 65°C with 1 x SSPE/0.1% SDS; if appropriate, the last step was repeated again (20 x SSPE is 3 M NaCl; 0.2 M NaH<sub>2</sub>PO<sub>4</sub>; 0.5 M EDTA; pH 7.4).

The hybridization was subsequently analyzed with the aid of a phosphoimager (Molecular Imager FX, BioRad).

- 40 For example, PstI-cut total DNA from different plants which had been regenerated after transformation with pCB199-3 were hybridized with the aadA gene as radiolabeled probe (793 bp PstI/NcoI fragment from pCB199-3). Here, it was found that the lines CB199NTH-4, -6 and -8 had indeed incorporated the aadA gene into the DNA. Moreover, EcoRI- and XhoI-cut total DNA from CB199NTH-4, -6 and -8 was hybridized with a radiolabeled probe (1082 bp BspI20I/SacI fragment from pCB199-3), which hybridizes

## 94

with part of the 16S rDNA. While, as expected, an approximately 3100 bp band was detected in the wild type (untransformed plant), mostly a band at 1750 bp was detected in the transplastomic lines, as the result of the insertion of the insertion sequence 5 from pCB199-3 into the plastome (Fig. 9). The resulting plants can be considered as being being predominantly homotransplastomic.

Example 5: Generation of transformation vectors which can be used 10 for transforming the plastids of master plants CB199NTH by means of the artificial homing process

## 5.1 Cloning of the homing endonuclease I-PpoI

15 The homing endonuclease I-PpoI was generated from 26 synthetic oligonucleotides by means of PCR, following a modification of the method of Stemmer WPC et al. (1995) Gene 164: 49-53 (SEQ ID NO: 11). The basic sequence was derived from the published sequence (Accession No. M38131 nucleotides 86 to 577). Here, a 20 few mutations were introduced to remove restriction endonuclease recognition sites from the gene; however, these mutations did not involve an altered amino acid sequence. The following elements were subsequently combined one after the other in a pBluescript KS (Stratagene) vector backbone in order to generate an I-PpoI 25 expression cassette. The sequence is flanked by the cleavage sites KpnI and SacI.

- a) Position 21 to 111: Prrn promoter
- 30 b) Position 118 to 135: 5'-untranslated region of the rbcL gene followed by 18 bp encoding 6 amino acids of the rbcL protein (bp 136-152)
- c) Position 154 to 645: Nucleic acid sequence encoding I-PpoI.
- 35 d) Position 688-779: 3'-untranslated region of the psbA gene.

The resulting plasmid was named pCB289-13. Despite the expression of the enzyme I-PpoI, which was expected to take place in E. 40 coli, no adverse effects on the growth were observed. The sequence described by SEQ ID NO: 4 resulted from the KpnI cleavage site to the SacI cleavage site (vector backbone remains that of pBluescript KS).

45 5.2 Generation of a transformation vector for artificial homing with homologous regions flanking the insertion sequence bilaterally

## 95

I) without I-PpoI in the insertion sequence

Regions around the I-PpoI recognition region from pCB199-3 were excised using PstI and SacI and ligated into the PstI and SacI cleavage sites of pBluescript. Thereafter, cleavage sites which were not required were removed from this vector by linearizing it with PstI and Bsp120I and, following treatment with Klenow fragment, recircularizing the vector. With the aid of commercially available enzyme I-PpoI (PROMEGA GmbH, Mannheim, Germany), the corresponding recognition region was cleaved in the resulting vector, and further cleavage sites were inserted therein by means of the synthetic oligos p190 and p191.

Oligo p190: 5'-GTCGACAGATCTTTAA-3' (SEQ ID NO: 20)

Oligo p191: 5'-AGATCTGTCGACTTAA-3' (SEQ ID NO: 21)

An expression cassette consisting of the following elements was introduced in the form of a BglII / XhoI fragment (SEQ ID NO: 6) into the cleavage sites SalI and BglII, which had thus been introduced:

- a) Prps16 promoter (complementary to 1033-1139)
- b) 5'rbcl (complementary to 1007-1024) with 18 bp encoding the 6 AAs (complementary to 989-1006)
- c) nptII gene (complementary to 185-988)
- d) 3'rbcl (complementary to 6-143)

The resulting vector was named pCB304-25 and also conferred kanamycin resistance to E. coli cells. This vector is no longer linearized by commercially available I-PpoI. All of the insert of the vector pCB304-25 (backbone pBluescript; replacing the MCS accordingly from SacI to KpnI) is described by SEQ ID NO: 62 and thus comprises the following elements:

- a) Position bp 19 to 110: 3'psbA from tobacco
- b) Position bp 149 to 160: nonfunctional "half" of the I-PpoI recognition sequence
- c) Position bp 171 to 277: Prps16 promoter

96

- d) Position bp 286 to 303: 5'rbcl sequence followed by 18 bp encoding the first 6 amino acids of the rbcL protein (bp 304-321)
- 5 e) Position bp 322 to 1125: nptII
- f) Position bp 1167 to 1304: 3'rbcl
- 10 g) Position bp 1310 to 1319: nonfunctional "half" of the I-PpoI recognition sequence

II) With I-PpoI in the insertion sequence

- 15 A BglIII / MunI fragment which encoded a 5'psbA - I-PpoI fusion was additionally introduced into the vector pCB304-25 with the aid of the BamHI and EcoRI cleavage sites. The resulting vector pCB320-192 thus expressed the nptII gene and I-PpoI homing endonuclease under the control of the Prps16 promoter. The
- 20 Bgl II / Mun I fragment is represented by SEQ ID NO: 63 and comprises the following elements:

- a) Position bp 6 to 82: 5'psbA
- 25 b) Position bp 83 to 574: I-PpoI

5.3 Generation of a transformation vector for artificial homing with a homologous region flanking the insertion sequence unilaterally

- 30 The elements located upstream of the Prps16 promoter and which are homologous to those in the master plants CB199NTH were removed from the vector pCB320-192 by restriction with KpnI and BglIII. Instead, a BstXI cleavage site was introduced therein by
- 35 means of synthetic oligonucleotides p199 and p200.

p199 5'-GATCTCCAGTTAACTGGGGTAC-3' (SEQ ID NO: 22)

p200 5'-CCCAGTTAACTGGA-3' (SEQ ID NO: 23)

- 40 DNA ends which are compatible with those originating by restriction with I-PpoI can now be generated by cleaving with BstXI. The resulting vector was renamed pCB322-1.
- A fragment which, at its one side, has an end which is compatible
- 45 with DNA which had been cleaved with I-PpoI at its core recognition region and, at its other side, homology with plastome

## 97

sequences of the master plants CB199NTH can be obtained from this vector for example using the enzymes BstXI and SacI.

5.4 Generation of a transformation vector for artificial homing  
5 without homologous regions around the insertion sequence

The remaining portion, which is homologous with recombinant plastid sequences of the master plants CB199NTH, was removed from the vector pCB322-1 using SacI and BspTI. Simultaneously, a BstXI  
10 cleavage site which, after cleavage with BstXI generates DNA ends which are compatible with I-PpoI-cut DNA, was generated here by introducing synthetic oligonucleotides p218 and p219. The resulting vector was named pCB347-33.

15 p218 5'-TTAAGCCAGTTAACTGGGCGGAGCT-3' (SEQ ID NO: 24)

p219 5'-CCGCCCAGTTAACTGGC-3' (SEQ ID NO: 25)

A fragment with bilateral DNA ends which are compatible with the  
20 DNA ends generated by the I-PpoI enzyme at its core recognition region can be isolated from this vector using the enzyme BstXI.

Example 6: Use of the master plants CB199NTH for plastid  
transformation by means of DSB induction

25

6.1 Using the vectors generated in 5.1 and 5.2 for transforming  
the master plants CB199NTH by using the DSB I enzyme I-PpoI

The plasmid pCB304-24 and the plasmid pCB289-13 were  
30 simultaneously applied to gold particles as described in Example 4 and used to bombard explants of the master plants CB199NTH, which explants had been treated analogously to what has been said in Example 4. However, the procedure differed from the description in Example 4 insofar as incubation is first carried  
35 out for 10 days on the regeneration medium without antibiotics; later, kanamycin is used in a concentration of 50 mg/l (in contrast to the 500 mg/l spectinomycin stated in Example 4).

The plasmid pCB320-192 was applied to gold particles as described  
40 in Example 4. After the ethanol washing step, 20 U of commercially available I-PpoI enzyme (Promega) were additionally added. Further treatment was as described above.

Also, in a different batch, 0.5 µg of a transcript generated in  
45 vitro with the aid of the T7 polymerase was applied simultaneously with the plasmid pCB320-192 to the gold particles. The template for the in-vitro transcription was

## 98

HindIII-linearized DNA of the plasmid pCB289-13. The transcript generated thus therefore encodes I-PpoI. After the bombardment, the treatment of the explants of the master plants continues as described above.

5

6.2 Using the vectors generated in 5.1 and 5.3 for transforming the master plants CB199NTH by utilizing the DSBI enzyme I-PpoI and homologous regions which are only unilaterally present

10

A fragment excised from the plasmid pCB322-1 with BstXI and SacI was eluted from an agarose gel. This fragment was subsequently applied to gold particles simultaneously with 1 µg of in-vitro transcript of HindIII-linearized plasmid pCB289-13 (cf. Example 15 6.1). After the explants of the master plants CB199NTH have been bombarded, the rest of the treatment is as described for Example 6.1.

6.3 Using the vectors generated in 5.1 and 5.4 for transforming 20 the master plants CB199NTH by utilizing the DSBI enzyme I-PpoI without homologous regions

The insertion sequence was excised from the plasmid pCB347-33 by means of BstXI and eluted from an agarose gel. This fragment was 25 applied to gold particles simultaneously with 1 µg of in-vitro transcript of the HindIII-linearized plasmid pCB289-13. The bombardment and the rest of the treatment are as detailed in Example 6.1.

30 Example 7: Identification of naturally occurring, endogenous recognition regions for homing endonucleases in plastomes of different plant species

Although no homing endonucleases are known to occur in the 35 plastids of higher plants, known plastome sequences were tested for the presence of recognition regions for homing endonucleases. This was done with the aid of the computer program SeqMan II (DNASTAR Inc.). The recognition regions which were identified in this manner are compiled in Table 1.

40

Based on the computer analysis, it was not possible to tell whether I-SceI has a recognition region in the plastidic genome or not. The region which is most likely to be able to act as recognition region was generated synthetically and integrated 45 into the XbaI and XhoI cleavage site of pBluescript in the form of oligonucleotides p276 and p277. The resulting plasmid pCB414-1 was subsequently analyzed with the aid of a commercially

## 99

available enzyme I-SceI (Roche) for the presence of a functional cleavage site. The plasmid was indeed linearized by I-SceI. This leads to the conclusion that I-SceI which is expressed in plastids likewise recognizes, and cleaves, this sequence. A further endogenous DSB recognition sequence for a DSB1 enzyme has thus been identified.

p276 5'-TCGAGAAGATCAGCCTGTTATCCCTAGAGTAACT-3' (SEQ ID NO: 26)

10 p277 5'-CTAGAGTTACTCTAGGGATAACAGGCTGATCTTC-3' (SEQ ID NO: 27)

Example 8: Cloning homologous regions from the tobacco plastome which flank the endogenous recognition region for the homing endonuclease I-CpaI

15

DNA fragments from the 23S rDNA of the tobacco plastome were amplified by means of PCR upstream and downstream of the I-CpaI recognition region using the primers p93 and p97, and p98 and p95, respectively.

20

p93: AAAGATCTCCTCACAAAGGGGGTCTG (SEQ ID NO: 64)

p97: TCGAAGACTTAGGACCGTTATAG (SEQ ID NO: 65)

25 p98: AGGAAGACCTTGTCGGGTAAGTTCCG (SEQ ID NO: 66)

p95: CTCAATTGGGGTCTCTCTGTCCAGGTGCAGG (SEQ ID NO: 67)

The resulting fragments were used for constructing the vector  
30 pCB270-1. The fragment from BssHII to BssHII of the pBlueScript vector (SEQ ID NO: 8) is shown, the 5' end of the sequence indicated being located at the BssHII cleavage site which is closer to the 3' end of the lacZ gene.

35 Two BpiI cleavage sites were introduced between the DNA fragments located upstream and downstream of the I-CpaI recognition regions. BpiI generates overhangs which are outside their recognition region. This procedure ensured that the vector pCB270-1 could likewise be used for the subsequent integration of  
40 various introns. To this end, simple overhangs which are compatible with the ends generated by BpiI in the vector pCB270-1 are generated at the introns to be cloned. Moreover, the respective nucleotides which are absent between the two fragments of the 23S rDNA in the vector pCB270-1 are added onto the  
45 introns. The selected regions are so highly conserved that there is no need to amplify new regions from other plant species. Furthermore, a point mutation in the 23S rDNA, as has also been



## 100

found in lincomycin-resistant mutants, has been introduced into the sequence downstream of the I-CpaI recognition region via PCR strategy. The sequence of the vector pCB270-1 which has been inserted into the pBluescript vector is shown in SEQ ID NO: 8.

5 The sequence comprises the following elements:

- Fragment of the 23S rDNA upstream of the I-CpaI recognition region: nucleotides 37 to 194
- 10 - Fragment of the 23SrDNA upstream of the I-CpaI recognition region: nucleotides 237 to 359
- Point mutation for lincomycin resistance: 352 (A being native at this point)

15

Vector pCB234-1 is constructed just as vector pCB270-1, but additionally has a recognition region for each of the restriction endonucleases XhoI and SacI downstream of the sequence shown hereinbelow.

20

Example 9: Cloning of the CpLSU2 intron including the homing endonuclease I-CpaI

The CpLSU2 intron (SEQ ID NO: 15) was amplified from the DNA of  
25 the alga *Chlamydomonas pallidostigmatica* (Culture Collection of Algae at the University of Göttingen, SAG Number 9.83, *Chlamydomonas segnis*, authentic strain of *Chlamydomonas pallidostigmatica* King) by means of PCR using the oligonucleotides p91 and p92.

30

p91 5'-AGAAGACGATCCTAAGG-3' (SEQ ID NO: 28)

p92 5'-TGAAGACTTGACAAGGAATTCGC-3' (SEQ ID NO: 29)

35 The oligonucleotides were chosen in such a way that cloning into the BpiI cleavage sites of the vector pCB234-1 was possible, as described above. The sequence comprises the following elements:

- Position 9-17 - portion of the tobacco 23S rDNA which is  
40 absent in pCB234-1
- CpLSU2 intron: position 18-893
- I-CpaI ORF: position 377-835

45

## 101

- Position 894-909 - portion of the tobacco 23S rDNA which is absent in pCB234-1

This fragment, which comprises the CpLSU2 intron, was cloned into the backbone of the vector pGEMTeasy (Promega) (vector pCB141-3). The entire fragment was excised from this vector using BpiI and cloned into the BpiI-linearized vector pCB234-1. The resulting vector was named pCB254-2.

# 10 Example 10: Nuclear LSU-rRNA intron from Tetrahymena thermophila

## 10.1 Cloning the LSU-rRNA intron from Tetrahymena thermophila

The LSU-rRNA intron was amplified from the organism Tetrahymena thermophila by means of PCR. Again, the oligonucleotides p102 and p103 were chosen in such a way that the nucleotides of the tobacco 23S rDNA, which are absent in pCB234-1, were added onto the intron to be amplified.

### 20 p102 (SEQ ID NO: 30):

5'-AGAAGACGATCCTAAATAGCAATATTTACCTTTGGGACCAAAAGTTATCAGGCATG-3'

### p103 (SEQ ID NO: 31):

5'TGAAGACTTGACAAGGAATTTTCGCTACCTTCGAGTACTCCAAACTAATC-3'

25

Moreover, the internal guide sequence (underlined in p102) is mutated in such a way over the wild type, owing to the choice of the oligonucleotide p102, that splicing of this intron at the desired position in the tobacco 23S rDNA is possible. The sequence shown in SEQ ID NO: 7 - the PCR fragment from the BpiI to BpiI cleavage site is shown - was cloned into the backbone of the vector pGEMTeasy. The resulting vector was named pCB220-17. The sequence comprises the following elements:

- 35 - Position 9-12 - portion of the tobacco 23S rDNA which is absent in pCB234-1
- LSU intron: position 13-425
- 40 - Position 426-446 - portion of the tobacco 23S rDNA which is absent in pCB234-1

The Tetrahymena LSU intron including the added, flanking sequences, was excised from the vector pCB220-17 using BpiI and inserted into the BpiI cleavage sites of the vector pCB234-1. The resulting product was named pCB255-1.

## 102

10.2 Indirect detection of the splicing activity of the  
Tetrahymena LSU intron in *E. coli*

To prove indirectly that the modified intron is indeed capable of  
5 splicing in the predetermined environment within the I-CpaI  
cleavage site, the modified Tetrahymena intron from pCB220-17  
together with a portion which surrounds the I-CpaI recognition  
region from the tobacco 23S rDNA was cloned in such a way into  
the lacZ gene of pBluescript that, if this intron is spliced into  
10 *E. coli* (strain XL1-Blue), a functional lacZ peptide can be  
formed. The expression of the latter can be detected in suitable  
strains by methods with which the skilled worker is familiar by  
converting the substance  
5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside (X-Gal) in the  
15 medium into a blue pigment. This vector was named pCB315-1. The  
lacZ gene including the introns in the vector pCB315-1 is  
described by SEQ ID NO: 9. The vector backbone is identical with  
pBluescript. The sequence comprises the following elements:

- 20 - lacZ-5' portion: complementary (789-765)
  - multiple cloning site from pBluescript: complementary  
(764-692)
- 25 - 23S rDNA fragment upstream and including the I-CpaI  
recognition region: complementary (691-682)
  - modified Tetrahymena intron: complementary (681-269)
- 30 - 23S rDNA fragment upstream and including the I-CpaI  
recognition region: complementary (268-244)
  - multiple cloning site from pBluescript: complementary  
(243-168)
- 35 - lacZ-5' portion: complementary (167-1)

A plasmid which corresponds to pCB315-1, but which plasmid  
(pCB305-1) lacks the element for the modified Tetrahymena intron,  
40 was generated for control purposes. pCB305-1 thus acted as  
positive control to demonstrate that lacZ, with the tobacco  
plastome 23S rDNA nucleotides incorporated in the reading frame  
is still functional. This reflects the situation after correct  
splicing of the Tetrahymena intron. XL1-Blue competent cells were  
45 transformed with the plasmids pCB315-1 and pCB305-1 by means of a  
method with which the skilled worker is familiar. In each case  
one individual colony was incubated on LB (Bactotryptone: 10 g/l,

## 103

yeast extract: 5 g/l, NaCl: 10 g/l, pH 7.5) plates comprising 15 g/l Bacto agar, 40 µg/l ampicillin, 75 µg/l IPTG (isopropyl-β-thiogalactopyranoside) and 80 µg/l X-Gal overnight at 37°C. In fact, both clones turned blue, which suggests that the modified Tetrahymena intron was spliced in the non-natural environment of the tobacco 23S rDNA in the heterologous organismus E. coli.

10.3: Introduction of further sequences into the Tetrahymena LSU intron

In addition to the experiments in Example 10.2, the possibility of incorporating further elements into the modified Tetrahymena intron without destroying the splicing activity was studied. To this end, pCB315-1 was linearized with BglII and the overhangs were filled up with the aid of Klenow fragment. Then, an XhoI-SacI fragment as is found in pCB199-3 was therefore cloned into this vector, likewise after treatment with Klenow fragment. A 229 bp fragment was thus inserted into the modified intron by this cloning step. This fragment comprises an I-PpoI recognition region. Independently of the orientation in which the 229 bp fragment inserted into the Tetrahymena intron, a blue coloration was detected in the test as described in Example 10.2. This suggests that the Tetrahymena intron is capable both of splicing at the desired region in the 23S rDNA and of incorporating additional genetic information while nevertheless retaining a splicing activity.

10.4 Transformation of a natural master plant and destruction of the endogenous I-CpaI recognition region with the modified Tetrahymena intron

The modified Tetrahymena intron was excised from the vector pCB220-17 using BpiI and cloned into the BpiI linearized vector pCB234-1 as described in principle in Example 8. The resulting vector was named pCB255-1.

pCB255-1 is applied to gold particles simultaneously with in vitro transcript of pCB262-5 (linearized with SalI, using T7 polymerase) by the method described in Example 4. These gold particles are subsequently used to bombard tobacco plants cv. Petit Havana analogously to the method described in Example 4. If appropriate, the explants can be selected on lincomycin (250 to 500 mg/l).

## 104

Example 11: L1.LtrB intron from *Lactococcus lactis*

The L1.LtrB intron including few bases of the flanking exon sequences was amplified from *Lactococcus lactis* by means of PCR using the primers p207 and p208. The PCR product was cloned into the vector pCR2.1TA (Invitrogen) (pCB345-34) and sequenced (SEQ ID NO: 10). Few deviations in comparison with the published sequence were found.

10 p207 5'-GAGAAGACATTCCTAACACATCCATAACGTGCG-3' (SEQ ID NO: 32)

p208 5'-TGAAGACTTGACATTGATATGGTGAAGTAGG-3' (SEQ ID NO: 33)

The cloned fragment in pCB345-34 (from the EcoRI cleavage site to the EcoRI cleavage site of the pCR2.1TA vector) is represented in SEQ ID NO: 10. The remainder of the vector is identical with the backbone of pCR2.1TA. The sequence comprises the following elements:

- 20 - Portion of the natural 5' exon: complementary (2540-2527)
- Intron L1.LtrB: complementary (2526-35)
- ORF in the intron: complementary (1953-154)
- 25 - Portion of the natural 3' exon: complementary (34-28)

Example 12: Generation of a further derivative of the *Tetrahymena* LSU intron, and incorporation of a foreign gene into this intron derivative, and transformation into natural master plants

In accordance with a preferred embodiment of the present invention, an artificial intron was generated, which intron can be incorporated into the plastidic genome at precisely the position where the natural intron belonging to the DSB recognition sequence under investigation resides. In the present example, the *Tetrahymena* LSU intron was modified in such a way that it is capable of splicing at the position marked ^^ at the recognition site, identified within the scope of the present invention, for the DSB enzyme I-CpaI in the plastidic genome of higher plants: CGATCCTAAGGT^AGCGAAATTCA.

The gene encoding I-CpaI including an RBS was subsequently incorporated into the intron. This gave rise to an intron which has splicing activity, bears a foreign gene and which can be

## 105

incorporated, by means of the process found within the scope of the present invention, into the plastids of a natural master plant within an essential gene (encoding the 23S rRNA).

5 12.1: Generation of a further Tetrahymena LSU intron derivative

To obtain a functional intron derivative at a predefined insertion site, the *internal guide sequence* (IGS) must be adapted in such a way that it is capable of undergoing base pairing with  
10 the 5' and the 3' exon. Fig. 10 illustrates how this adaptation was carried out in the present example in order to generate a Tetrahymena intron derivative which is capable of splicing at the natural insertion site of the CpLSU5 intron within the I-CpaI recognition region. An adaptation to any desired insertion site  
15 can be carried out analogously. The intron generated within the scope of the present example was named TetIVS2a and is described by SEQ ID NO: 73.

20 12.2: Indirect detection of the splicing activity of the TetIVS2a intron in *E. coli*

TetIVS2a was incorporated into the lacZ gene of pBluescript analogously to Example 10.2. After suitable incubation of *E. coli* XL1-blue cells which comprised the plasmid pCB459-1, a blue  
25 coloration indicated the splicing activity of the TetIVS2a intron at the desired position.

Components of the insert from plasmid pCB459-1 (SEQ ID NO: 74; backbone corresponds to pBluescript)

- 30
- lacZ-3' portion including parts of the multiple cloning site from pBluescript (complementary to position 1-254)
  - Sequence from the I-CpaI recognition region (complementary to position 254-265)
  - 35 - TetIVS2a (complementary to position 266-678)
  - Sequence from the I-CpaI recognition region (complementary to position 679-687)
  - lacZ-5' portion including parts of the multiple cloning site from pBluescript (complementary to position 688-791)
- 40

12.3: Introduction of further genetic information into the TetIVS2a intron and detection of the splicing activity in *E. coli*

45 In this example, the gene encoding the DSB1 enzyme I-CpaI is introduced into TetIVS2a without the latter losing its splicing activity at said position within the I-CpaI recognition region.

## 106

To this end, a BclI cleavage site was first introduced, by PCR, into the sequence segment of TetIVS2a which corresponds to loop L8 in the Tetrahymena LSU intron. A nonfunctional derivative of the gene encoding I-CpaI was then incorporated into this BclI  
5 cleavage site. Since the expression of I-CpaI in E. coli is toxic, it was necessary to use, for the splice test in E.coli, a nonfunctional gene which had previously been generated by linearizing the gene in question at the EcoRI cleavage site, making the overhangs blunt-ended by Klenow treatment and  
10 subsequently religating the gene segments. This resulted in a reading-frame shift in the gene. Incorporation of said intron with the nonfunctional gene into the lacZ gene of pBluescript gave rise to the plasmid pCB478-3 and, again, it was possible to detect the splicing activity of this intron in E. coli at the  
15 desired position within the I-CpaI recognition site by means of the blue coloration of colonies in question, analogously to Example 12.2. Since the functional gene encoding I-CpaI differs from the nonfunctional gene used in pCB478-3 by only 4 bases, it can be assumed that the intron retains the desired splicing  
20 activity, even after the functional, instead of the nonfunctional, I-CpaI gene has been incorporated.

Components of the insert of plasmid pCB478-3 (SEQ ID NO: 75; backbone corresponds to pBluescript)

- 25
- lacZ-3' portion including parts of the multiple cloning site from pBluescript (complementary to position 1-265)
  - Sequence from the I-CpaI recognition region (complementary to position 256-265)
  - 30 - TetIVS2a (complementary to position 266-1178), comprising a nonfunctional gene for I-CpaI (complementary to position 399-861) and an RSB upstream of the nonfunctional I-CpaI gene (complementary to 866-870)
  - Sequence from the I-CpaI recognition region (complementary to  
35 position 1179-1187)
  - lacZ-5' portion including parts of the multiple cloning site from pBluescript (complementary to position 1179-1291)

40 12.4: Transformation of a self-disseminating, artificial intron in a natural master plant

After it had been demonstrated that the TetIVS2a intron is capable of splicing at the desired position and of simultaneously incorporating further genetic information without losing this  
45 splicing activity, a construct was generated which is intended to make possible that the I-CpaI gene can be introduced into the plastidic genome by means of the method described within the

## 107

scope of the present invention in the form of a foreign gene without using a selection marker. To this end, the vector pCB492-25, which comprises an insert with the following elements was first generated (SEQ ID NO: 76; backbone corresponds to that of pBluescript; sequence from BssHII to BssHII in pBluescript is indicated, the BssHII cleavage site indicated here at the 5' end is the BssHII cleavage site in pBluescript which is localized closer to the 3' end of the lacZ gene):

- 10 - 23S rDNA fragment upstream of and including the I-CpaI recognition region (position 37-203)
- TetIVS2a (position 204-1112) with inserted gene encoding I-CpaI (position 521-979) and RBS (position 512-516)
- 23S rDNA fragment downstream of and including the I-CpaI recognition region (position 1113-1247)

To ensure expression of I-CpaI directly after the introduction into plastids of natural master plants, a promoter was added in vitro upstream of said intron Cpa construct by means of PCR. The primers p396 and p95 and, as template, pCB492-25 were used for this purpose.

p396 (SEQ ID NO: 77):

5'-TAGTAAATGACAATTTTCTCTGAATTATATAATTAACATGGCGACTGTTTACCAAAAAC-3'

p95 (SEQ ID NO: 78): 5'-CTCAATTGGGGTCTCTCTGTCCAGGTGCAGG-3'

The resulting PCR product was named Prom-TetIVS2a-Cpa, is described by SEQ ID NO: 79 and comprised the following elements:

- 30 - synthetic promoter (position 8-40)
- tobacco 23S rDNA upstream of and including portions of the I-CpaI recognition region (position 41-207)
- TetIVS2a (position 208-1116) comprising gene encoding I-CpaI (position 525-983) and RBS (position 516-520)
- 35 - tobacco 23S rDNA downstream of and including portions of the I-CpaI recognition region (position 1117-1243)

The plasmid pCB492-25 was applied to gold particles simultaneously with the above-described PCR product Prom-TetIVS2a-Cpa as described in Example 4 and subsequently used to bombard wild-type tobacco. By expression of the I-CpaI enzyme, it was intended to bring about a double-strand break in the 23S rDNA, which double-strand break is repaired by the PCR product which has been introduced or by the insertion sequence, of the plasmid pCB492-25, which has been introduced. The I-CpaI recognition region, which is naturally present, is thereby



## 108

inactivated in the transformed plastome copies. Plants were regenerated without any selection pressure, and these plants are tested by PCR for the presence of the insertion sequence in the plastome.

5

Example 13: Dissemination of the modified Tetrahymena LSU intron from pCB255-1 in a natural master plant by expression of the DSBI enzyme I-CpaI *in trans*

10 This example shows how a DSBI enzyme can be expressed in the plastids of master plants in order to efficiently disseminate an insertion sequence in the copies of the master plant.

13.1: Generation of a vector for the transformation of plastids  
15 which permits the expression of the homing endonuclease I-CpaI in plastids

First, a vector which encodes the selection marker *aadA* and the DSBI enzyme I-CpaI was generated. Since expression of the I-CpaI  
20 enzyme is lethal in *E. coli*, the *accD* promoter was chosen in order to allow for the expression of this enzyme in the plastids, but to prevent the expression in *E. coli*. Thus, it was possible to generate and amplify this vector in a conventional manner with *E. coli* as the host organism. The resulting vector was named  
25 pCB435-45 and comprised an insert as shown in SEQ ID NO: 80 with the following elements:

- Right-hand target region (as in pCB42-94, see above; complementary to position 66-1403)
- 30 - promoter *PaccD* (position 1422-1478)
- RBS (position 1500-1504)
- Gene encoding I-CpaI (position 1509-1967)
- Expression cassette for the marker gene *aadA* consisting of:
  - the 3' region of the *psbA* gene (complementary to position  
35 2065-1974)
  - *aadA* gene (complementary to position 2872-2078)
  - 5'-untranslated regions of the tobacco *rbcL* gene (complementary to position 2890-2873), partly mutated
- Promoter of the gene for the 16S rRNA (complementary to  
40 position 2987 to 2897)
- Left-hand target region (as in pCB42-94, see above; complementary to position 3863-3007)
- Portions of the pBluescript (including origin of replication; positions 3864-4746 and 1-65)

45

## 109

## 13.2 Cotransformation of pCB435-45 and pCB255-1 into natural master plants

The plasmids pCB435-45 and pCB255-1 were applied simultaneously to gold particles as detailed in Example 4 and then introduced into plastids of tobacco leaves by means of the particle gun. Transplastomic plants were selected on regeneration medium supplemented with 500 mg/l spectinomycin as described in Example 4. As soon as plantlets had formed, they were transferred to growth medium supplemented with 500 mg/l spectinomycin, and leaf material was harvested. This leaf material was analyzed by Southern analysis using the Dig-Easy Hyb® (Roche Diagnostics; Mannheim) for the incorporation of the two plasmids into the plastidic genome. A probe with a sequence as shown in SEQ ID NO: 81 was used to determine the percentage of plastome copies which were transgenic with regard to the insertion sequence from pCB435-45 (probe directed against portions of the 16S rDNA).

A probe with a sequence as shown in SEQ ID NO: 82 was used to determine the percentage of plastome copies which were transgenic with regard to the insertion sequence from pCB255-1 (probe directed against portions of the 23S rDNA).

Fig. 11 shows that, in this experiment, 2 lines (CB255+435NTH-19 and -20) were identified which are transgenic with regard to both the insertion sequence of pCB435-45 and that from pCB255-1. It was furthermore demonstrated in this manner that, surprisingly, the insertion sequence from pCB255-1 (modified Tetrahymena LSU intron) had already been disseminated into more copies of the plastidic genome than the insertion sequence from pCB435-45, even though the selection had been carried for the event of the insertion of the insertion sequence from pCB435-45 (aadA marker gene resides in pCB435-45). The efficiency of the method described within the present invention - viz. the insertion and rapid dissemination of an insertion sequence in the plastidic genome without selecting for the presence of this insertion sequence by utilizing DSBI enzymes and suitable recognition sites - has thus been demonstrated for said lines in the present example.

40

45

## 110

Example 14: Generation of further master plants with a DSB recognition region which does not naturally occur in plastids, and transformation of these plants utilizing the DSB1 enzyme I-PpoI

5

14.1: Generation of a further vector (pCB456-2) for introducing a non-naturally-occurring recognition region for the homing endonuclease I-PpoI into the plastome of tobacco

- 10 The purpose of this approach was (analogously to Example 3) to generate a further vector for the transformation of plastids, which vector has no extensive homologies with sequences in the plastidic genome.
- 15 In this plasmid, the selection marker *aadA* is under the control of a synthetic promoter which is derived from the consensus sequence for *E. coli*  $\sigma 70$  promoters. A region downstream of the *Synechocystis* 3'psbA-1 gene was used as the 3' end. In contrast to the vector pCB199-3 which has already been described, the DSB
- 20 recognition sequence was here introduced into the molecule immediately downstream of the *aadA* gene, but upstream of the *Synechocystis* 3'psbA-1 sequence. An operon can be generated thereby with the aid of a DSB1 enzyme following insertion of an insertion sequence. The genes on the insertion sequence can then
- 25 be optionally inserted on the insertion sequence without promoter. After the insertion, suitable genes of the insertion sequence then also come under the control of the synthetic promoter upstream of the *aadA* gene in the master plant. An operon structure consisting of the *aadA* and the subsequently introduced
- 30 genes can thereby optionally be generated in the plastome.

Various elements were cloned one after the other into the basic vector pCB42-94 in order to generate the plasmid pCB456-2:

- 35 - Synthetic promoter (complementary to bp 1226-1260)  
- Ribosome binding site (complementary to bp 1214-1218)  
- *aadA* gene (complementary to bp 414-1208)  
- Core recognition region for the homing endonuclease I-PpoI (complementary to bp 331-345)
- 40 - 3'psbA-1 from *Synechocystis* (complementary to bp 19-155)

The vector thus obtained also confers spectinomycin resistance in *E. coli*. This vector, which is named pCB456-2, comprises the abovementioned elements within the nucleic acid sequence with the

45 SEQ ID NO: 83, instead of the multiple cloning site in the basic

## 111

vector pCB42-94 for the transformation of plastids. Again, all of the sequence which replaces MCS (from SacI to KpnI) is indicated.

- 14.2: Generation of predominantly homotransplastomic master plants which comprise a nonnatural DSB recognition sequence

The vector pCB456-2 was introduced into the plastids of tobacco analogously to pCB199-3 in Example 4. However, as opposed to the description in Example 4, the shoots obtained were grown on growth medium comprising 30 g/l sucrose (instead of the 10 g/l stated in Example 4). The resulting plants were named CB456NTH. 2 lines which have the insertion sequence from pCB456-2 incorporated into their plastome (CB456NTH-1 and -15, cf. Fig. 12) were identified among the spectinomycin-resistant plants obtained after the transformation, using Southern hybridization. A probe which was directed against a fragment of the 16S rDNA was employed in the Southern experiment (cf. Example 13.2 above). This probe was suitable for detecting an approx. 3.1 kb fragment from EcoRI-digested DNA corresponding to the wild type. In contrast, an approx. 1.7 kb fragment was detected when the insertion sequence from pCB456-2 had been incorporated into the corresponding plastome copies.

- 14.3: Generation of a transformation vector for artificial homing in the master plants CB456NTH

First, an operon structure consisting of the elements RBS - nptII (encoding an enzyme which confers kanamycin resistance) - RBS - I-PpoI (encoding a DSB1 enzyme) was generated. This cassette was surrounded with BstXI cleavage sites which, after exposure to the enzyme BstXI, generate DNA ends which are compatible with the DNA ends generated by the enzyme I-PpoI. The resulting vector (backbone corresponds to that of pBluescript) was named pCB528-2 and comprises an insert as shown in SEQ ID NO: 84 with the following elements:

- RBS (position 28-32)
- nptII (position 27-840)
- RBS (position 849-853)
- Gene encoding I-PpoI (position 859-1350)

The 1360 bp fragment was subsequently excised from pCB528-2 using BstXI and ligated into the I-PpoI cleavage site in the vector pCB456-2. Clones with kanamycin resistance were selected from those obtained after the ligation products had been transformed into E. coli. It was thereby ensured that said insert in the

## 112

clone in question was inserted in the vector in such a way that the nptII and I-PpoI cassettes had the same orientation as the aadA cassette. This was also verified by the restriction analysis method, with which the skilled worker is familiar. The vector in question was named pCB535-11.

14.4: Transformation of pCB535-11 into master plants CB456NTH  
pCB535-11 was applied to gold particles as described for pCB456-2 in Example 14.2 and subsequently introduced into plastids of the  
10 master plant CB456NTH-1 using the particle gun.

Some of the explants were incubated on regeneration medium without any selection pressure. Resulting plants were transferred to growth medium (again without selection pressure). Thereafter,  
15 the plants are analyzed by PCR for the presence of the RBS - nptII - RBS - I-PpoI cassette.

Other explants were incubated on regeneration medium supplemented with 15 or 30 mg/l kanamycin. After 2 weeks, the plants are  
20 transferred to fresh regeneration medium and the kanamycin concentration increased stepwise to 50 and 80 mg/l, respectively.

25

30

35

40

45

0817/00028

1

## SEQUENCE LISTING

<110> SunGene GmbH & Co.KGaA  
 <120> Method for the transformation of **vegetal plastids**  
 <130> AE20010854 / PF 0081700028  
 <140>  
 <141>  
 <160> 84  
 <170> PatentIn Ver. 2.1  
 <210> 1  
 <211> 4363  
 <212> DNA  
 <213> Artificial sequence  
 <220>  
 <223> Description of the artificial sequence: plastid  
 transformation plasmid pCB42-94  
 <220>  
 <221> misc\_feature  
 <222> (55)..(1405)  
 <223> right targeting region  
 <220>  
 <221> mutation  
 <222> (346)  
 <223> mutation causing streptomycin resistance  
 <220>  
 <221> mutation  
 <222> (68)  
 <223> mutation causing spectinomycin resistance  
 <220>  
 <221> misc\_feature  
 <222> (1404)..(1511)  
 <223> multiple cloning site  
 <220>  
 <221> misc\_feature  
 <222> (2629)..(3417)  
 <223> Ampicillin resistance  
 <400> 1  
 gcgccaata cgcaaaccgc ctctcccgc gcgttgccg attcattaat gcagtatga 60  
 cttgacggca tctcacctt cctccggctt atcaccggca gtctgttcag ggttccaaac 120  
 tcaacgatgg caactaaaca cgagggttgc gctcgttgcg ggacttaacc caacacctta 180  
 cggcagcagc tgacgacagc catgcaccac ctgtgtccgc gttcccgaag gcaccctct 240  
 ctttcaagag gattcgcggc atgtcaagcc ctggttaagg tcttcgctt gcatcgaatt 300

0817/00028

2

aaaccacatg ctccaocgct tgtcggggcc ccggtcaatt ccttttagtt tcattcttgc 360  
 gaacgtactc cccaggcggg atacttaacg cgtagctac agcactgcac gggctgatac 420  
 gcacagcgcc tagtatccat cgtttacggc taggactact ggggtatcta atccattcg 480  
 ctcccctagc ttctgtctct cagtgtcagt gtcggcccag cagagtgtct tcgccgttgg 540  
 tgttctttcc gatctctacg catttcaccg ctccaccgga aattccctct gccctaccg 600  
 tactccagct tggtagtttc caccgcctgt ccagggttga gccctgggat ttgacggcgg 660  
 acttaaaaag ccacctacag acgctttacg cccaatcatt ccggataacg ctgcatcct 720  
 ctgtattacc gcggctgctg gcacagagtt agccgatgct tattccccag ataccgtcat 780  
 tgccttctct ccgggaaaag aagttcacga ccggtgggcc ttctacctcc acgcggcatt 840  
 gctocgtcag gctttcgccc attgcggaaa attccccact gctgcctccc gtaggagtct 900  
 gggccgtgtc tcagtccag tgggctgat catcctctcg gaccagctac tgatcatcgc 960  
 cttggttaagc tattgcctca ccaactagct aatcacagcg gagccccctc tcgggcggat 1020  
 tctctcttt gctctcagc ctacggggtg ttgacagccg ttccagctg ttgtccct 1080  
 cccaagggca ggttcttacg cgttactcac ccgtccgcca ctggaaacac cacttccgt 1140  
 ccgacttga tgtgtaagc atccgccag cgttcatcct gagccaggat cgaactctcc 1200  
 atgagattca tagttgatt acttatagct tcttgttcg tagacaaagc ggattcggaa 1260  
 ttgtcttca ttccaaggca taactgtat ccatgcgct catattgcc cggagtgcg 1320  
 tccagaaat atagccatcc ctgccccctc acgtcaatcc cagagcctc ttatccattc 1380  
 tcattgaacg acggcggggg agcgagctcc accgcggtgg cggccgctct agaactagt 1440  
 gatcccccg gctgcaggaa ttgatatca agcttatcga taccgtcgac ctgaggggg 1500  
 ggccccgtac caaatccaac tagaaaaact cacattgggc ttagggataa tcaggctcga 1560  
 actgatgact tccaccagct caagtgaca cttaccgct gagttatc cttccccgc 1620  
 cccatcgaga aatagaactg actaatccta agtcaaaggg tcgagaaact caacgccact 1680  
 attctgaac aactggagc cgggcctct ttgcacta ttacggatat gaaaataatg 1740  
 gtcaaaatcg gattcaattg tcaactgcc ctatcggaag taggattgac taccgattcc 1800  
 gaaggaaactg gagttacatc tctttccat tcaagagttc ttatgcgtt ccacgccct 1860  
 ttgagacccc gaaaaatgga caaatcctt ttctaggaa cacatacaag attcgtcact 1920  
 acaaaaagga taatggaac cctaccatta actattcat ttatgaatt catagtaata 1980  
 gaaatacatg tctaccgag acagaattg gaactgcta tctcttccc tagcaggcaa 2040  
 agatttacct ccgtggaaag gatgattcat tcggatcgac atgagagtc aactacattg 2100  
 ccagaatcca tgtgtatat ttgaaagagg ttgacctct tcttctctc atgttact 2160  
 cctctcccg ccgagcccc ttctctctg gtccacagag acaaatgta ggactgggtc 2220  
 caacaattca tcagactcac taagtggga tcaactaata atactaatct aatataatg 2280  
 tctaataat ctaataat agaaaatact aatataatg aaaagaactg tctttctgt 2340  
 atactttccc cggttccgtt gctaccgagg gcctcgtgat acgcctatt ttataggta 2400  
 atgcatgat aataatggtt tcttagacgt cagggtggcac tttcgggga aatgtgcgcg 2460  
 gaacctctat ttgttatt ttctaaatac attcaaatat gtatccgctc atgagacaat 2520  
 aacctgata aatgcttcaa taatatgaa aaaggaagag tatgagtatt caacatttcc 2580  
 gtgtgccct tattccctt ttgcggcat ttgccttcc tgttttgc caccagaaa 2640  
 cgctggtgaa agtaaaagat gctgaagatc agttgggtgc acgagtgggt tacatcgaac 2700  
 tggatctcaa cagcggtgaa atccttgaga gttttcgccc cgaagaactg ttccaatga 2760  
 tgagcactt taaagtctg ctatgtggcg cgttattat ccgtattgac gccgggcaag 2820  
 agcaactcgg tcgcccata cactattctc agaatgactt ggttgagtac tcaccagtca 2880  
 cagaaaagca tcttacggat ggcatgacag taagagaatt atgcagtgc gccataacca 2940  
 tgagtataa cactgcggcc aactlactc tgacaacgat cggaggaccg aaggagctaa 3000  
 ccgcttttt gcacaacatg ggggatcatg taactgcct tgatcgttgg gaaccggagc 3060

0817/00028

3

tgaatgaagc cataccaaac gacgagcgtg acaccacgat gcctgtagca atggcaacaa 3120  
 cgttgcgcaa actattaact ggcgaactac ttactctagc ttcccggcaa caattaatag 3180  
 actggatgga ggcggataaa gttgcaggac cacttctcgc ctccggccct cccgctggct 3240  
 ggtttattgc tgataaatct ggagccgggtg agcgtgggtc tcgcgggtatc attgcagcac 3300  
 tggggccaga tggtaagccc tccgtatcg tagttatcta cagcagggg agtcaggcaa 3360  
 ctatggatga acgaaataga cagatcgtg agatagggtc ctactgatt aagcattgt 3420  
 aactgtcaga ccaagtttac tcatatatac tttagattga tttaaaactt cattttaat 3480  
 ttaaaaggat ctaggtaag atccttttg ataattcat gaccaaatac ccttaacgtg 3540  
 agttttcgtt ccactgagcg tcagaccccg tagaaaagat caaaggatct tcttgagatc 3600  
 cttttttct gcgcgtaac tgctgctgc aaacaaaaa accaccgcta ccagcgggtg 3660  
 ttgtttgcc ggatcaagag ctaccaactc ttttccgaa ggtaactggc tcagcagag 3720  
 cgcagatacc aaatactgic ctctagtggt agccgtagtt aggccaccac ttcaagaact 3780  
 ctgtagcacc gcctacatac ctgcctcgc taatcctgtt accagtggct gctgccagt 3840  
 gcgataagtc gtgtcttacc gggttggact caagacgata gttaccgat aaggcgcagc 3900  
 ggtcgggctg aacgggggggt tcgtgcacac agcccagctt ggagcgaacg acctacaccg 3960  
 aactgagata cctacagcgt gagctatgag aaagcggcac gcttccgaa gggagaaagg 4020  
 cggacaggta tccggttaagc ggcaggggtc gaacaggaga gcgcagagg gagcttcag 4080  
 ggggaaacgc ctggtatctt tatagtcctg tcgggttctg ccacctctga cttgagcgtc 4140  
 gattttgtg atgctcgtca gggggcgga gcctatgaa aaagccagc aacgcggcct 4200  
 tttacgggt cctggcctt tgctggcct ttgtcacat gttcttct cgttatccc 4260  
 ctgattctgt ggataaccgt attaccgct ttgagtgcg tgataccgt cgcgcagcc 4320  
 gaacgaccga gcgcagcag tcagtgcg aggaagcga aga 4363

&lt;210&gt; 2

&lt;211&gt; 1359

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence: Insert of  
 vector pCB199-3

&lt;400&gt; 2

gagctctgat cacggaagat agctttggca aaaaaagcaa aaagcattta ccttgattga 60  
 gatgttaatt gtgttgcaa ttatcagtat ttiaattttg cttttgtgc caaattgat 120  
 actagagctt cgggtgccag ggcgtgccct tgggtcccc gggcgcgtac tcgacgctac 180  
 cttagagag tcaagcttct atattaccct gttatcccta gcgtactcga gaaaaaaga 240  
 aaggagcaat agcacctct ttagagaaca agaaaatgat taitgtcct tcttttcaa 300  
 aacctctat agactaggcc aggaattatc tgcagttat tgccaactac cttagtgatc 360  
 tcgctttca cgtagtggac aaattcttcc aactgatct gcgcgagggc caagcgtatc 420  
 tcttctgtc caagataagc ctgtctagct tcaagtatga cgggtgata ctgggccggc 480  
 aggcgctcca ttgccagtc ggcagcgaca tcttcggcg cgattttgcc ggttactgcg 540  
 ctgtacaaa tgcgggacaa cgtaagcact acatttcgt catcgccagc ccagtgggc 600  
 ggcgagttcc atagcgttaa ggttcattt agcgctcaa atagatcctg ttcaggaacc 660  
 ggatcaaaaga gttctccgc cgtggacct accaaggcaa cgctatgtc tctgtcttt 720  
 gtcagcaaga tagccagatc aatgtcgtc gtggctggct cgaagatacc tgcaagaatg 780  
 tcattgcgt gccatttcc aaattgcagt tcgcgttag ctggataacg ccacggaatg 840  
 atgtcgtct gcacaacaat ggtgacttct acagcggga gaatctcgt cttccaggg 900  
 gaagccgaag ttccaaaag gtcgttgatc aaagctgcc cgttgtttc atcaagcctt 960



0817/00028

4

acgggtcaccg taaccagcaa atcaatatca ctgtgtggct tcaggccgcc atccactgcg 1020  
 gagccgtaca aatgtacggc cagcaacgtc ggttcgagat ggcgctcgat gacgccaaact 1080  
 acctctgata gttgagttga tacttcggcg ataaccgctt cagcagccat ggctgtttgt 1140  
 ggtgtcatgg ctgtttgtgg tgtcatgaat cctccctac aactagatcc tcgcccggag 1200  
 ttgctccca gaaatatagc catccctgcc cctcagtc aatccacga gctcttctc 1260  
 cattctcatt gaacgacggc gtcgaggggg ggcccgggtac gtcgacgaag ttcctattcc 1320  
 gaagttccta tictcaagaa agtataggaa ctctgtacc 1359

&lt;210&gt; 3

&lt;211&gt; 1070

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence: insert of  
 vector pCB401-20

&lt;400&gt; 3

gagctctgat caggaagat agctttggca aaaaaagcaa aaagcattta ccttgattga 60  
 gatgttaatt gtgtggcaa ttatcagtat ttaattttg cttttgtgc caaatttgat 120  
 actagagctt cgggtgccag ggcgtgccct tgggctcccc gggcgcgtac tcgacgctac 180  
 cttagagag tcaagcttct atattaccct gttatcccta gcgtactcga gctgcagtta 240  
 ttgccaaact accttagtga tctcgcttt cagtagtgg acaaattctt ccaactgac 300  
 tgcgcgcgag gccaagcga cttctcttg tccaagataa gcctgtctag ctcaagtat 360  
 gacgggctga tactgggccc gcaggcgtc cattgcccag tcggcagcga catccttcgg 420  
 cgcgattttg ccggttacg cgtgtacca aatgcgggac aacgtaagca ctacatttcg 480  
 ctatcgcca gccagtcgg gcggcgagtt ccatagcgtt aaggtttcat ttacgcctc 540  
 aaatagatcc tgttcaggaa ccggatcaaa gatttctcc gccgctggac ctaccaaggc 600  
 aacgctatgt tctcttgctt ttgtcagcaa gatagccaga tcaatgtcga tcgtggctgg 660  
 ctgaagata cctgcaagaa tgcattgcg ctgccattct ccaaattgca gttcgcgtt 720  
 agctggataa cgccacggaa tgatgtcgtc gtgcacaaca atggtgactt ctacagcgcg 780  
 gagaatctcg ctctctccag gggaagccga agtttccaaa aggtcgttga tcaaagctcg 840  
 ccgcgttgtt tcatcaagcc ttacggtcac cgtaaccagc aaatcaatat cactgtgttg 900  
 cttcaggccg ccattccatg cggagccgta caaatgtacg gccagcaacg tcggttcgag 960  
 atggcgtcgt atgacgcaa ctaccttga tagttgagtt gatacttcgg cgataaccgc 1020  
 ttacagagcc atggttcct cctacaacg tcgagggggg gcccggtacc 1070

&lt;210&gt; 4

&lt;211&gt; 804

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence: insert of  
 vector pCB289-13

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (154)..(642)

&lt;223&gt; sequence coding for I-PpoI homing-endonuclease

0817/00028

5

&lt;400&gt; 4

ggtagcgggc cccccctcga cgccgtcgtt caatgagaat ggataagagg ctctgtggat 60  
 tgacgtgagg gggcagggat ggctatattt ctgggagcga actccgggag aggatctagt 120  
 tgtagggagg gattcatgac accacaaaca gcc atg gcg ctc acc aat gct caa 174  
 Met Ala Leu Thr Asn Ala Gln

1 5

atc ttg gct gtg att gac agt tgg gaa gaa aca gtc ggt cag ttt cca 222  
 Ile Leu Ala Val Ile Asp Ser Trp Glu Glu Thr Val Gly Gln Phe Pro  
 10 15 20

gtg ata acg cac cat gta cca tta ggt ggc ggt ctg caa gga acg ctc 270  
 Val Ile Thr His His Val Pro Leu Gly Gly Gly Leu Gln Gly Thr Leu  
 25 30 35

cat tgt tac gag atc ccc cta gca gct cct tat ggg gtt ggc ttt gct 318  
 His Cys Tyr Glu Ile Pro Leu Ala Ala Pro Tyr Gly Val Gly Phe Ala  
 40 45 50 55

aag aat ggg cct acc cgc tgg caa tac aaa cgg aca atc aat caa gtc 366  
 Lys Asn Gly Pro Thr Arg Trp Gln Tyr Lys Arg Thr Ile Asn Gln Val  
 60 65 70

gtc cac aga tgg ggg tcc cac aca gtc cct ttt cta tta gaa ccg gat 414  
 Val His Arg Trp Gly Ser His Thr Val Pro Phe Leu Leu Glu Pro Asp  
 75 80 85

aac atc aac ggc aaa acc tgc aca gca tgc cac cta tgt cat aat act 462  
 Asn Ile Asn Gly Lys Thr Cys Thr Ala Ser His Leu Cys His Asn Thr  
 90 95 100

cga tgc cac aat ccc ttg cac ttg tgc tgg gag tca cta gac gac aac 510  
 Arg Cys His Asn Pro Leu His Leu Cys Trp Glu Ser Leu Asp Asp Asn  
 105 110 115

aaa ggc aga aac tgg tgc ccc ggt ccc aac ggg gga tgt gtc cat gcg 558  
 Lys Gly Arg Asn Trp Cys Pro Gly Pro Asn Gly Gly Cys Val His Ala  
 120 125 130 135

gtg gtt tgt tta agg cag ggt ccg ttg tac ggc cca ggg gcg act gtg 606  
 Val Val Cys Leu Arg Gln Gly Pro Leu Tyr Gly Pro Gly Ala Thr Val  
 140 145 150

gca ggt cct caa caa agg ggc agt cac ttt gtg gta taactgcaga 652  
 Ala Gly Pro Gln Gln Arg Gly Ser His Phe Val Val  
 155 160

agcttcaatt gcatgctcta gatgatcaaa gaattcctgg cctagtctat aggaggtttt 712

gaaaagaaag gagcaataat cattttcttg ttctatcaag aggggtgctat tgctccttc 772

ttttttctc gagaggggta cgtaccgagc tc 804

&lt;210&gt; 5

&lt;211&gt; 163

0817/00028

6

&lt;212&gt; PRT

&lt;213&gt; Artificial sequence

<223> Description of the artificial sequence: insert of  
vector pCB289-13

&lt;400&gt; 5

Met Ala Leu Thr Asn Ala Gln Ile Leu Ala Val Ile Asp Ser Trp Glu

1 5 10 15

Glu Thr Val Gly Gln Phe Pro Val Ile Thr His His Val Pro Leu Gly

20 25 30

Gly Gly Leu Gln Gly Thr Leu His Cys Tyr Glu Ile Pro Leu Ala Ala

35 40 45

Pro Tyr Gly Val Gly Phe Ala Lys Asn Gly Pro Thr Arg Trp Gln Tyr

50 55 60

Lys Arg Thr Ile Asn Gln Val Val His Arg Trp Gly Ser His Thr Val

65 70 75 80

Pro Phe Leu Leu Glu Pro Asp Asn Ile Asn Gly Lys Thr Cys Thr Ala

85 90 95

Ser His Leu Cys His Asn Thr Arg Cys His Asn Pro Leu His Leu Cys

100 105 110

Trp Glu Ser Leu Asp Asp Asn Lys Gly Arg Asn Trp Cys Pro Gly Pro

115 120 125

Asn Gly Gly Cys Val His Ala Val Val Cys Leu Arg Gln Gly Pro Leu

130 135 140

Tyr Gly Pro Gly Ala Thr Val Ala Gly Pro Gln Gln Arg Gly Ser His

145 150 155 160

Phe Val Val

&lt;210&gt; 6

&lt;211&gt; 1150

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

&lt;223&gt; Description of the artificial sequence:

XhoI / Bgl II fragment for construction of  
vector pCB304-25

&lt;400&gt; 6

ctcagagtatt cggctcaatc ctttagtaa aagattgggc cgagtttaatt tgcaattcaa 60

taaagagaac gaaggataat tacttgagtt cttctcctt atcctcttt atttctgct 120

aatttaictg ctaatgtcta ctgaattcaa gcttgatcc gcggccgcct agcttggttc 180

ccgctcagaa gaactcgtca agaaggcgat agaaggcgat gcgctgcgaa tcgggagcgg 240

0817/00028

7

cgataccgta aagcacgagg aagcggtag cccattcgcc gccaaagtct tcagcaatat 300  
cacgggtagc caacgctatg tcctgatagc ggtagccac accagccgg ccacagtga 360  
tgaatccaga aaagcggcca ttccacca tgataatgg caagcaggca tcgcatggg 420  
tcacgacgag atcctcgccg tcgggcatgc gcgccttgag cctggcgaac agttcggtg 480  
gcgcgagccc ctgatgctct tegtccagat catctgac gacaagaccg gttccatcc 540  
gagtagctgc tcgtcgatg cgatgttcg ctgggtggc gaatgggcag gtagccgat 600  
caagcgtatg cagccggcg attgcatcag ccatgatgga tactttctg gcaggagcaa 660  
ggtgagatga caggagatcc tgccccggca ctccgcca tagcagccag tccctccc 720  
cttcagtac aacgtcgagc acagctgcgc aaggaaacgcc cgtcgtggc agccacgata 780  
gccgcgtgc ctgcctgc agttcattca gggcaccgga caggtcggc ttgacaaaa 840  
gaaccgggcg cccctgcgt gacagccgga acacggcggc atcagagcag ccgattgtct 900  
gttggtccca gtcataccg aatagcctct ccaccaagc ggccggagaa cctgcgtgca 960  
atccatctg ttcaatcaa gtcctatgg ctgttggtg tgcataat cctccctac 1020  
aactagatcc tatatactat agagataggt ggataaatat ttcttttag taagaccca 1080  
tcgctaatat taattatct aacatattaa ttaatattha atatataat atatatagag 1140  
tcgagatcta 1150

&lt;210&gt; 7

&lt;211&gt; 454

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence: insert of  
vector pCB220-17

&lt;400&gt; 7

gaagacgac cttaatagca atattacct ttgggaccaa aagtatcag gcatgcaact 60  
ggtagctagt cttaaacca atagattgca tcggtttaa aggaagacc gtcaaattgc 120  
gggaaagggg tcaacagccg ttcatgacca agtctcagg gaaacttga gatggcctg 180  
caaagggtat ggtataaagc tgacggacat ggtcttaacc acgcagccaa gtcttaagtc 240  
aacagatctt ctgtgatat ggatgcagt cacagactaa atgtcggtcg gggaagatgt 300  
attcttctca taagatatag tcggacctct ccttaatggg agctagcgga tgaagtatg 360  
caacactgga gccgctggga actaattgt atgcgaaagt atattgatta gtttggagt 420  
actcgaaggt agcgaaattc ctgtcaagt ctgc 454

&lt;210&gt; 8

&lt;211&gt; 498

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence: insert of  
vector pCB270-1

&lt;400&gt; 8

gcgcgcgtaa tacgactcac tataggcgga attgggggcg actgtttacc aaaaacacag 60  
gtctccgcaa agtcgtaaga ccatgtatgg gggctgacgc ctgccagtg ccggaaggtc 120  
aaggaggttg gtgacctgat gacaggggag ccggcgaccg aagccccggg gaacggcggc 180  
cgtaactata acggtcctaa gtctcgaat cgaattcact agtgatagga agacctgtc 240  
gggtaagtic cgaccgcac gaaaggcgta acgatctggg cactgtctcg gagagaggct 300

0817/00028

8

cgggtgaata gacatgtctg tgaagatgcg gactacctgc acctggacag agagacccca 360  
 attcctggcc tagtctatag gaggttttga aaagaaagga gcaataatca ttttctgtt 420  
 ctatcaagag ggtgctattg ctctttctt ttttctcga ccagcttttg ttcccttag 480  
 tgagggttaa ttgcgcgc 498

&lt;210&gt; 9

&lt;211&gt; 789

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence: lacZ gene  
 with inserted intron to demonstrate splicing  
 (from vector pCB315-1)

&lt;400&gt; 9

acaatttcca ttgccattc aggtgctgca actgttggga agggcgatcg gtgcgggcct 60  
 ctctgctatt acgccagctg gcgaaagggg gatgtgctgc aaggcgatta agttgggtaa 120  
 cgccagggtt ttccagtcga cgacgttgta aaacgacggc cagtgaagcg gcgtaatacg 180  
 actcactata gggcgaattg ggtaccgggc cccccctcga ggtcgacggg atcgataagc 240  
 ttgacccgac aaggaatttc gctaccttcg agtactccaa aactaatcaa tatactttcg 300  
 catacaaatt agttcccagc ggctccagtg ttgcatcact tcatccgcta gctcccatta 360  
 aggagaggtc cgactataac ttatgagaag aatacatctt ccccgaccga catttagtct 420  
 gtgaactgca tccatatcaa cagaagatct gttgacttag gacttggctg cgtggtagg 480  
 accatgtccg tcagcttatt accataccct ttgcaaggcc atctcaaagt ttccctgag 540  
 acttggctact gaacggctgt tgacccctt cccgcaattt gacggcttg cttttaaac 600  
 cgatgcaatc tattggttta aagactagct accagggtgca tgctgataa cttttgtcc 660  
 caaaggtaaa tattgctatt taggaccgtt aggatccact agttctagag cggccgccac 720  
 cgcgggtggag ctccagctt ttgtccctt agtgagggtt aattgcgcgc ttggcgtaat 780  
 catggtcat 789

&lt;210&gt; 10

&lt;211&gt; 2566

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence: LI.LtrB  
 intron from vector pCB345-34.

&lt;400&gt; 10

gaattcgccc ttgaagact tgacatttga tatggtgaag tagggaggta ccgccttgtt 60  
 cacattactg tgactgggtt gcaccacct ctccgggaac cgtacgtacc cctctcggag 120  
 tatacggctc tgttattgtt cgctcgtaaa aattcacttg tgttatgaa tcacgtgacg 180  
 atgacaatga aagcatacaa caagagttt acgtgtttc gctatcattg ccatttccca 240  
 ttttctttg cctttaagat tttagacctt attgacatgg tgaattcat aggaagtatt 300  
 ttcatcagat gtccacata attcacaaca tttagctttt aacctgtttt caagagtatt 360  
 ccgggcatag ccatacaata caggagcttg acttaicta tccgtaaatt gataagggga 420  
 ttacattca ctaaaattg caaaalaacg gcgctgctta ccttgcctta tctcatacgg 480  
 gatgccccac gaaccacitc catctttaa catggaaatg gttttgaaa gtgtccctt 540  
 atgtttggag gctatcggtt tttagacagt gtattccata agataagcaa aataattgag 600

0817/00028

9

ctggttaaaa ttacttgcta gaccgtagta attacaaatc cctcttaatt cagaattata 660  
 aattgtgatg atttctaagt ctgttgaacg aataagatat ttctgtgtaa ctggaacca 720  
 tgagctatct ttcttttga tagctatctt ctgtcaaaa ataaattgac gaattttgtc 780  
 ttgaagagga ataaggaggt ctacactccc attgagtggt ctcttttga ctttaccaga 840  
 tcgttttacc gtccacttc tcttactcg tatatcatat ccagaaaaac gagcgggttg 900  
 actgctatgt gtgatgagtg tttttcttc actcaattcc atttttagct tgttatgaat 960  
 aaaaaagttt aattgttctt ttatccattg acagtccctt ttgcttcctt taacagagat 1020  
 aatgaagtcg tccgcatacc ggacgtattt caatacttta ttgtctgtg aggtacaggg 1080  
 gagtgggggt aatcttttac gttttcttg atattctaaa agaacttag cttttcttc 1140  
 accctccaac ttcttgagac ggtgagaaat tcttttacc tcattgtgaa gttcccgata 1200  
 ttccagtggt attcttctg gactttctg gtcaaaactt attttgaggt gtaaaacaaa 1260  
 ctatccaat tcatgaagat agatgttggc caaaagagga gatagaattc cacttgagg 1320  
 tgttccgctg taagtttgt gatactgcca gttttccaga taacctgctt itagaaattt 1380  
 ataaatcaat tggctcattt tcatatctt gattttaaga ttgatgagtc caatgagtg 1440  
 aacgtggctt atattatcga agcagccctt tatatctccc tccacaaacc atctgccc 1500  
 gccaaactct cttttgattg tticaaagc cgtgtgacag ctctgttgag gtctaaaacc 1560  
 gtgagacaca tcttcgaata ccggttcata gatagattca agaattatc tcacagcttc 1620  
 ttggaatcaat ttatctgta aagttggaat tccaaagggt ctatctttt tagaattctt 1680  
 ttitgcaata tacattcttc gtacaggttg aggatagtaa gttccgtctt ttaaagattg 1740  
 aataatctt ttattttt cticactaaa gccatccgct gtatcatcta atattcctt 1800  
 tgtggaagct cttttattgg aatataaatt ttgatacgcc acgtaataaa tatctggagc 1860  
 taaaagataa cgataaagtc ttgtaaaaac ttctctata ttcttcttg aattttact 1920  
 gattcttctt aaaattgcc a ttgttggtt cattttgagg ttctctccc taatcaattt 1980  
 ttaattttag tacacaataa ctgtaccctt ttgccatgta aaggcggtta ccttctcag 2040  
 actactacga gtactccgta ccttgcaag atttcaagc tctagtgtga tagcctttt 2100  
 cctcttctt attaggcatt ctgtttagg gtatccccag ttagtgtta gtcttggtta 2160  
 attcagatc tcggcatcgc ttctgttcg ttcccatagg ttctctaca gattgtaca 2220  
 atgtgggat aacagataag tccacaacca taactaacct ttcttgtag tagaggttc 2280  
 agacacttc ctctatcgac acataaccga aattagaaac ttgckttcag taaacacaac 2340  
 ttatacctta tatctgatta acattgcgac tcagtcgtac ccgattgtct ttaggttaact 2400  
 catcgcttc caaccgtgct ctgtcccggt atcagcttc gctttcgggt taggttggt 2460  
 gttttctgt ttatctaca gagtagtacc taaactact tgacttaaca ccctatctgg 2520  
 gcgcacgtta tggatgtgtt aggaatgtct tctcaagggc gaattc 2566

&lt;210&gt; 11

&lt;211&gt; 520

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence: synthetic  
 sequence coding for I-Ppo I homing-endonuclease

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (16)..(504)

&lt;400&gt; 11

gcatatgaga tctcc atg gcg ctc acc aat gci caa atc ttg gct gtg att 51

0817/00028

10

Met Ala Leu Thr Asn Ala Gln Ile Leu Ala Val Ile  
 1 5 10

gac agt tgg gaa gaa aca gtc ggt cag ttt cca gtg ata acg cac cat 99  
 Asp Ser Trp Glu Glu Thr Val Gly Gln Phe Pro Val Ile Thr His His  
 15 20 25

gta cca tta ggt ggc ggt ctg caa gga acg ctc cat tgt tac gag atc 147  
 Val Pro Leu Gly Gly Gly Leu Gln Gly Thr Leu His Cys Tyr Glu Ile  
 30 35 40

ccc cta gca gct cct tat ggg gtt ggc ttt gct aag aat ggg cct acc 195  
 Pro Leu Ala Ala Pro Tyr Gly Val Gly Phe Ala Lys Asn Gly Pro Thr  
 45 50 55 60

cgc tgg caa tac aaa cgg aca atc aat caa gtc gtc cac aga tgg ggg 243  
 Arg Trp Gln Tyr Lys Arg Thr Ile Asn Gln Val Val His Arg Trp Gly  
 65 70 75

tcc cac aca gtc cct ttt cta tta gaa ccg gat aac atc aac ggc aaa 291  
 Ser His Thr Val Pro Phe Leu Leu Glu Pro Asp Asn Ile Asn Gly Lys  
 80 85 90

acc tgc aca gca tgc cac cta tgt cat aat act cga tgc cac aat ccc 339  
 Thr Cys Thr Ala Ser His Leu Cys His Asn Thr Arg Cys His Asn Pro  
 95 100 105

ttg cac ttg tgc tgg gag tca cta gac gac aac aaa ggc aga aac tgg 387  
 Leu His Leu Cys Trp Glu Ser Leu Asp Asp Asn Lys Gly Arg Asn Trp  
 110 115 120

tgc ccc ggt ccc aac ggg gga tgt gtc cat gcg gtg gtt tgt tta agg 435  
 Cys Pro Gly Pro Asn Gly Gly Cys Val His Ala Val Val Cys Leu Arg  
 125 130 135 140

cag ggt ccg ttg tac ggc cca ggg gcg act gtg gca ggt cct caa caa 483  
 Gln Gly Pro Leu Tyr Gly Pro Gly Ala Thr Val Ala Gly Pro Gln Gln  
 145 150 155

agg ggc agt cac ttt gtg gta taactgcagc tcgagg 520  
 Arg Gly Ser His Phe Val Val  
 160

<210> 12  
 <211> 163  
 <212> PRT  
 <213> Artificial sequence  
 <223> Description of the artificial sequence: synthetic  
 sequence coding for I-Ppo I homing-endonuclease  
 <400> 12  
 Met Ala Leu Thr Asn Ala Gln Ile Leu Ala Val Ile Asp Ser Trp Glu  
 1 5 10 15  
 Glu Thr Val Gly Gln Phe Pro Val Ile Thr His His Val Pro Leu Gly  
 20 25 30

0817/00028

11

Gly Gly Leu Gln Gly Thr Leu His Cys Tyr Glu Ile Pro Leu Ala Ala  
 35 40 45  
 Pro Tyr Gly Val Gly Phe Ala Lys Asn Gly Pro Thr Arg Trp Gln Tyr  
 50 55 60  
 Lys Arg Thr Ile Asn Gln Val Val His Arg Trp Gly Ser His Thr Val  
 65 70 75 80  
 Pro Phe Leu Leu Glu Pro Asp Asn Ile Asn Gly Lys Thr Cys Thr Ala  
 85 90 95  
 Ser His Leu Cys His Asn Thr Arg Cys His Asn Pro Leu His Leu Cys  
 100 105 110  
 Trp Glu Ser Leu Asp Asp Asn Lys Gly Arg Asn Trp Cys Pro Gly Pro  
 115 120 125  
 Asn Gly Gly Cys Val His Ala Val Val Cys Leu Arg Gln Gly Pro Leu  
 130 135 140  
 Tyr Gly Pro Gly Ala Thr Val Ala Gly Pro Gln Gln Arg Gly Ser His  
 145 150 155 160  
 Phe Val Val

&lt;210&gt; 13

&lt;211&gt; 481

&lt;212&gt; DNA

&lt;213&gt; Chlamydomonas pallidostigmatica

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (4)..(459)

&lt;223&gt; ORF coding for I-Cpa I homing endonuclease

&lt;400&gt; 13

acc atg gac att aat cct caa tgg att aca ggt ttc gta gat ggg gaa 48

Met Asp Ile Asn Pro Gln Trp Ile Thr Gly Phe Val Asp Gly Glu

1 5 10 15

ggt tgt ttt agt gta agt ata ctt aga aat aat tcg ttg cgc tat ggc 96

Gly Cys Phe Ser Val Ser Ile Leu Arg Asn Asn Ser Leu Arg Tyr Gly

20 25 30

cat cag ctt caa cca gaa ttc gta gtg acc caa cat aaa tta gat gca 144

His Gln Leu Gln Pro Glu Phe Val Val Thr Gln His Lys Leu Asp Ala

35 40 45

aat gtt tta tat gca tta aaa gac tac ttt aaa gtt gga tca gtc gtt 192

Asn Val Leu Tyr Ala Leu Lys Asp Tyr Phe Lys Val Gly Ser Val Val

50 55 60

gtg aat cat ggg gaa cgg ctt tgc tat aaa gtc aaa aat att gat cac 240

Val Asn His Gly Glu Arg Leu Cys Tyr Lys Val Lys Asn Ile Asp His

65 70 75



0817/00028

12

ttt ata acc gtc att ata cca ttt ttc gaa aaa cat gag cta aaa aca 288  
 Phe Ile Thr Val Ile Ile Pro Phe Phe Glu Lys His Glu Leu Lys Thr  
 80 85 90 95  
 aaa aga aga att gaa ttt ctt cga ttt cga aaa atc tgc ttg ctg tta 336  
 Lys Arg Arg Ile Glu Phe Leu Arg Phe Arg Lys Ile Cys Leu Leu Leu  
 100 105 110  
 aaa gca ggt aga cat tta gaa tcg cag gaa gga ttc gag aaa gtg ttg 384  
 Lys Ala Gly Arg His Leu Glu Ser Gln Glu Gly Phe Glu Lys Val Leu  
 115 120 125  
 gat tta gca aaa aaa ctc cgt atc aat gag aaa aac tac cag gaa tct 432  
 Asp Leu Ala Lys Lys Leu Arg Ile Asn Glu Lys Asn Tyr Gln Glu Ser  
 130 135 140  
 atc aaa cgt ttt gaa gaa act ggc gag taactcgaga gtatagagct cc 481  
 Ile Lys Arg Phe Glu Glu Thr Gly Glu  
 145 150  
 <210> 14  
 <211> 152  
 <212> PRT  
 <213> Chlamydomonas pallidostigmatica  
 <400> 14  
 Met Asp Ile Asn Pro Gln Trp Ile Thr Gly Phe Val Asp Gly Glu Gly  
 1 5 10 15  
 Cys Phe Ser Val Ser Ile Leu Arg Asn Asn Ser Leu Arg Tyr Gly His  
 20 25 30  
 Gln Leu Gln Pro Glu Phe Val Val Thr Gln His Lys Leu Asp Ala Asn  
 35 40 45  
 Val Leu Tyr Ala Leu Lys Asp Tyr Phe Lys Val Gly Ser Val Val Val  
 50 55 60  
 Asn His Gly Glu Arg Leu Cys Tyr Lys Val Lys Asn Ile Asp His Phe  
 65 70 75 80  
 Ile Thr Val Ile Ile Pro Phe Phe Glu Lys His Glu Leu Lys Thr Lys  
 85 90 95  
 Arg Arg Ile Glu Phe Leu Arg Phe Arg Lys Ile Cys Leu Leu Leu Lys  
 100 105 110  
 Ala Gly Arg His Leu Glu Ser Gln Glu Gly Phe Glu Lys Val Leu Asp  
 115 120 125  
 Leu Ala Lys Lys Leu Arg Ile Asn Glu Lys Asn Tyr Gln Glu Ser Ile  
 130 135 140  
 Lys Arg Phe Glu Glu Thr Gly Glu  
 145 150

0817/00028

13

<210> 15  
 <211> 917  
 <212> DNA  
 <213> Chlamydomonas pallidostigmatica

<220>  
 <221> intron  
 <222> (18)..(893)  
 <223> CPLSU2 Intron

<400> 15  
 gaagacgac ctaagggttaa ttgctagctt tagtaaaact gaactatatg ctggaaaatc 60  
 ctccggccctc tctttccccc agttgcactt cgtgctaaag cgggcagatg aaactataaa 120  
 gcaaaaagag gctcatcact ttttttacc ggaacccgt tccggcttaa agtgatgaaa 180  
 aagatgtctt gtacgtctcg tattttgcca cgagccgtga taatcaagcg tacatgggga 240  
 caatcagcag ggaaggatgt ttgatacacg cagagtagaa ctggcagtct agaatcggtc 300  
 ggtatcgcta aacataaccc tcagagacta tacgttcgga agtcttaata atcataataa 360  
 tagggaggaaa ttgaaatgg acattaatcc tcaatggatt acaggtttcg tagatgggga 420  
 aggtgtttt agtgaagta tacttagaaa taatcgttg cgctatggcc atcagcttca 480  
 accagaattc gtagtgacc aacataaatt agatgcaaat gttttatatg cattaaaaga 540  
 ctactttaa gttggatcag tcgttgtgaa tcatggggaa cggcttgct ataaagtcaa 600  
 aaatattgat cactttataa ccgtcattat accattttc gaaaaacatg agctaaaaac 660  
 aaaaagaaga atgaatttc ttcgatttcg aaaaatctgc ttgctgttaa aagcaggtag 720  
 acatttagaa tcgcaggaag gattcgagaa agtgttggat tiagcaaaaa aactccgtat 780  
 caatgagaaa aactaccagg aatctatcaa acgttttgaa gaaactggcg agtaaaaaaa 840  
 ataagattta agatagatc cagcctattt tgaagaaaaa tggggtaaat ctgagcgaaa 900  
 ttcttgtca agtcttc 917

<210> 16  
 <211> 21  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> Description of the artificial sequence:  
 oligonucleotide primer

<400> 16  
 taaggccctc ggtagcaacg g 21

<210> 17  
 <211> 20  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> Description of the artificial sequence:  
 oligonucleotide primer

<400> 17  
 ggggtaccaa atccaactag 20

0817/00028

14

<210> 18  
 <211> 25  
 <212> DNA  
 <213> Artificial sequence  
 <220>  
 <223> Description of the artificial sequence:  
     oligonucleotide primer  
 <400> 18  
 ggagctcgct cccccgccgt cggtc 25  
 <210> 19  
 <211> 27  
 <212> DNA  
 <213> Artificial sequence  
 <220>  
 <223> Description of the artificial sequence:  
     oligonucleotide primer  
 <400> 19  
 gatgcatgat gacttgacgg catcctc 27  
 <210> 20  
 <211> 16  
 <212> DNA  
 <213> Artificial sequence  
 <220>  
 <223> Description of the artificial sequence:  
     oligonucleotide primer  
 <400> 20  
 gtcgacagat ctttaa 16  
 <210> 21  
 <211> 16  
 <212> DNA  
 <213> Artificial sequence  
 <220>  
 <223> Description of the artificial sequence:  
     oligonucleotide primer  
 <400> 21  
 agatctgtcg acttaa 16  
 <210> 22  
 <211> 22  
 <212> DNA  
 <213> Artificial sequence  
 <220>  
 <223> Description of the artificial sequence:  
     oligonucleotide primer

0817/00028

15

<400> 22  
 gatctccagt taactggggt ac 22  
 <210> 23  
 <211> 14  
 <212> DNA  
 <213> Artificial sequence  
 <220>  
 <223> Description of the artificial sequence:  
 oligonucleotide primer  
 <400> 23  
 cccagttaac tgga 14  
 <210> 24  
 <211> 25  
 <212> DNA  
 <213> Artificial sequence  
 <220>  
 <223> Description of the artificial sequence:  
 oligonucleotide primer  
 <400> 24  
 ttaagccagt taactgggcg gagct 25  
 <210> 25  
 <211> 17  
 <212> DNA  
 <213> Artificial sequence  
 <220>  
 <223> Description of the artificial sequence:  
 oligonucleotide primer  
 <400> 25  
 ccgccagtt aactggc 17  
 <210> 26  
 <211> 34  
 <212> DNA  
 <213> Artificial sequence  
 <220>  
 <223> Description of the artificial sequence:  
 oligonucleotide primer  
 <400> 26  
 tcgagaagat cagcctgtta tcctagagt aact 34  
 <210> 27  
 <211> 34  
 <212> DNA  
 <213> Artificial sequence

0817/00028

16

&lt;220&gt;

<223> Description of the artificial sequence:  
oligonucleotide primer

&lt;400&gt; 27

ctagagttac tctaggata acaggctgat cttc 34

&lt;210&gt; 28

&lt;211&gt; 17

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence:  
oligonucleotide primer

&lt;400&gt; 28

agaagacgat cctaagg 17

&lt;210&gt; 29

&lt;211&gt; 23

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence:  
oligonucleotide primer

&lt;400&gt; 29

tgaagacttg acaaggaatt tcg 23

&lt;210&gt; 30

&lt;211&gt; 56

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence:  
oligonucleotide primer

&lt;400&gt; 30

agaagacgat cctaaatagc aatatttacc ttgggacca aaagttatca ggcatg 56

&lt;210&gt; 31

&lt;211&gt; 50

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence:  
oligonucleotide primer

&lt;400&gt; 31

tgaagacttg acaaggaatt tcgctacctt cgagtactcc aaaactaatc 50

&lt;210&gt; 32

&lt;211&gt; 33

0817/00028

17

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence:  
oligonucleotide primer

&lt;400&gt; 32

gagaagacat tcctaacaca tccataacgt gcg 33

&lt;210&gt; 33

&lt;211&gt; 32

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence:  
oligonucleotide primer

&lt;400&gt; 33

tgaagacttg acatttgata tggatgaagta gg 32

&lt;210&gt; 34

&lt;211&gt; 175

&lt;212&gt; DNA

&lt;213&gt; Pisum sativum

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)..(174)

&lt;223&gt; plastid translocalization sequence

&lt;400&gt; 34

atg gct tct atg ata tcc tct tca gct gtg act aca gtc agc cgt gct 48  
Met Ala Ser Met Ile Ser Ser Ser Ala Val Thr Thr Val Ser Arg Ala  
1 5 10 15tct acg gtg caa tgg gcc gcg gtg gct cca ttc ggc ggc ctc aaa tcc 96  
Ser Thr Val Gln Ser Ala Ala Val Ala Pro Phe Gly Gly Leu Lys Ser  
20 25 30atg act gga ttc cca gtt aag aag gtc aac act gac att act tcc att 144  
Met Thr Gly Phe Pro Val Lys Lys Val Asn Thr Asp Ile Thr Ser Ile  
35 40 45aca agc aat ggt gga aga gta aag tgc atg c 175  
Thr Ser Asn Gly Gly Arg Val Lys Cys Met  
50 55

&lt;210&gt; 35

&lt;211&gt; 58

&lt;212&gt; PRT

&lt;213&gt; Pisum sativum

0817/00028

18

&lt;400&gt; 35

Met Ala Ser Met Ile Ser Ser Ser Ala Val Thr Thr Val Ser Arg Ala

1 5 10 15

Ser Thr Val Gln Ser Ala Ala Val Ala Pro Phe Gly Gly Leu Lys Ser

20 25 30

Met Thr Gly Phe Pro Val Lys Lys Val Asn Thr Asp Ile Thr Ser Ile

35 40 45

Thr Ser Asn Gly Gly Arg Val Lys Cys Met

50 55

&lt;210&gt; 36

&lt;211&gt; 84

&lt;212&gt; PRT

&lt;213&gt; Nicotiana tabacum

&lt;220&gt;

&lt;221&gt; TRANSIT

&lt;222&gt; (1)..(84)

<223> plastidic translocalisation signal derived from  
plastidic transketolase

&lt;400&gt; 36

Met Ala Ser Ser Ser Ser Leu Thr Leu Ser Gln Ala Ile Leu Ser Arg

1 5 10 15

Ser Val Pro Arg His Gly Ser Ala Ser Ser Ser Gln Leu Ser Pro Ser

20 25 30

Ser Leu Thr Phe Ser Gly Leu Lys Ser Asn Pro Asn Ile Thr Thr Ser

35 40 45

Arg Arg Arg Thr Pro Ser Ser Ala Ala Ala Ala Val Val Arg Ser

50 55 60

Pro Ala Ile Arg Ala Ser Ala Ala Thr Glu Thr Ile Glu Lys Thr Glu

65 70 75 80

Thr Ala Gly Ser

&lt;210&gt; 37

&lt;211&gt; 258

&lt;212&gt; DNA

&lt;213&gt; Nicotiana tabacum

&lt;220&gt;

&lt;221&gt; transit\_peptide

&lt;222&gt; (1)..(258)

<223> plastidic translocalisation signal derived from  
plastidic transketolase (frame 1)

0817/00028

19

&lt;400&gt; 37

ggtaccatgg cgtctcttc ttctctcact ctctctcaag ctatctctc tegtctgtc 60  
 cctcgccatg gctctgcctc ttctctcaa ctttccctt ctctctcac ttttccggc 120  
 cttaaacca atccaatat caccacctcc cgcgcgcta ctcttctc cgcgcgcc 180  
 gccgcgctg taaggcacc ggcgattcgt gccacagctg caaccgaaac catagagaaa 240  
 actgagactg cgggatcc 258

&lt;210&gt; 38

&lt;211&gt; 260

&lt;212&gt; DNA

&lt;213&gt; Nicotiana tabacum

&lt;220&gt;

&lt;221&gt; transit\_peptide

&lt;222&gt; (1)..(260)

<223> plastidic translocalisation signal derived from  
 plastidic transketolase (frame 2)

&lt;400&gt; 38

ggtaccatgg cgtctcttc ttctctcact ctctctcaag ctatctctc tegtctgtc 60  
 cctcgccatg gctctgcctc ttctctcaa ctttccctt ctctctcac ttttccggc 120  
 cttaaacca atccaatat caccacctcc cgcgcgcta ctcttctc cgcgcgcc 180  
 gccgcgctg taaggcacc ggcgattcgt gccacagctg caaccgaaac catagagaaa 240  
 actgagactg cgctgatcc 260

&lt;210&gt; 39

&lt;211&gt; 259

&lt;212&gt; DNA

&lt;213&gt; Nicotiana tabacum

&lt;220&gt;

&lt;221&gt; transit\_peptide

&lt;222&gt; (1)..(259)

<223> plastidic translocalisation signal derived from  
 plastidic transketolase (frame 3)

&lt;400&gt; 39

ggtaccatgg cgtctcttc ttctctcact ctctctcaag ctatctctc tegtctgtc 60  
 cctcgccatg gctctgcctc ttctctcaa ctttccctt ctctctcac ttttccggc 120  
 cttaaacca atccaatat caccacctcc cgcgcgcta ctcttctc cgcgcgcc 180  
 gccgcgctg taaggcacc ggcgattcgt gccacagctg caaccgaaac catagagaaa 240  
 actgagactg cggggatcc 259

&lt;210&gt; 40

&lt;211&gt; 63

&lt;212&gt; PRT

&lt;213&gt; Arabidopsis thaliana

&lt;220&gt;

&lt;221&gt; TRANSIT

&lt;222&gt; (1)..(63)

<223> plastidic translocalisation signal derived from



0817/00028

20

plastidic isopentenylpyrophosphate isomerase-2  
(IPP-2)

&lt;400&gt; 40

Met Ser Ala Ser Ser Leu Phe Asn Leu Pro Leu Ile Arg Leu Arg Ser  
1 5 10 15

Leu Ala Leu Ser Ser Ser Phe Ser Ser Phe Arg Phe Ala His Arg Pro  
20 25 30

Leu Ser Ser Ile Ser Pro Arg Lys Leu Pro Asn Phe Arg Ala Phe Ser  
35 40 45

Gly Thr Ala Met Thr Asp Thr Lys Asp Gly Ser Arg Val Asp Met  
50 55 60

&lt;210&gt; 41

&lt;211&gt; 205

&lt;212&gt; DNA

&lt;213&gt; Arabidopsis thaliana

&lt;220&gt;

&lt;221&gt; transit\_peptide

&lt;222&gt; (1)..(205)

<223> plastidic translocalisation signal derived from  
plastidic isopentenylpyrophosphate isomerase-2  
(IPP-2) (frame 1)

&lt;400&gt; 41

gataccaca ccaacaccaa tgcctgcttc ttctttattt aatctcccat tgattgcct 60  
cagatctctc gctctttcgt ctctttttc ttctttccga ttgcccatc gtctctgtc 120  
atcgatttca ccgagaaagt taccgaattt tgcctgcttc tctggatccg ctatgacaga 180  
tactaaagat ggatcccggtcgcac 205

&lt;210&gt; 42

&lt;211&gt; 207

&lt;212&gt; DNA

&lt;213&gt; Arabidopsis thaliana

&lt;220&gt;

&lt;221&gt; transit\_peptide

&lt;222&gt; (1)..(207)

<223> plastidic translocalisation signal derived from  
plastidic isopentenylpyrophosphate isomerase-2  
(IPP-2) (frame 2)

&lt;400&gt; 42

gataccaca ccaacaccaa tgcctgcttc ttctttattt aatctcccat tgattgcct 60  
cagatctctc gctctttcgt ctctttttc ttctttccga ttgcccatc gtctctgtc 120  
atcgatttca ccgagaaagt taccgaattt tgcctgcttc tctggatccg ctatgacaga 180  
tactaaagat ctggatcccg ggtcgcac 207

&lt;210&gt; 43

&lt;211&gt; 206

0817/00028

21

&lt;212&gt; DNA

&lt;213&gt; Arabidopsis thaliana

&lt;220&gt;

&lt;221&gt; transit\_peptide

&lt;222&gt; (1)..(206)

<223> plastidic translocalisation signal derived from  
plastidic isopentenylpyrophosphate isomerase-2  
(IPP-2) (frame 3)

&lt;400&gt; 43

gatatccaca ccaacaccaa tgtctgcttc ttctttatt aatctcccat tgattcgctt 60  
 cagatctctc gctcttcgt ctctttttc ttctttccga ttgcccac gtcctctgtc 120  
 atcgatttca ccgagaaagt taccgaattt tcgtgcttc tctggtaccg ctatgacaga 180  
 tactaaagat gggatcccgg gtcgac 206

&lt;210&gt; 44

&lt;211&gt; 44

&lt;212&gt; DNA

&lt;213&gt; Nicotiana tabacum

&lt;220&gt;

&lt;221&gt; promoter

&lt;222&gt; (1)..(44)

&lt;223&gt; plastidic promoter PrbcL

&lt;400&gt; 44

gttgcgctat atatatgaaa gagtatacaa taatgatgta ttg 44

&lt;210&gt; 45

&lt;211&gt; 64

&lt;212&gt; DNA

&lt;213&gt; Nicotiana tabacum

&lt;220&gt;

&lt;221&gt; promoter

&lt;222&gt; (1)..(64)

&lt;223&gt; plastidic promoter Prps16-107

&lt;400&gt; 45

tagcgatggg gtcttactaa agaaaaatat ttatccacct atctctatag tatatagata 60  
 taga 64

&lt;210&gt; 46

&lt;211&gt; 91

&lt;212&gt; DNA

&lt;213&gt; Nicotiana tabacum

&lt;220&gt;

&lt;221&gt; promoter

&lt;222&gt; (1)..(91)

&lt;223&gt; plastidic promoter Prn16

0817/00028

22

<400> 46  
 cgccgtcgtt caatgagaat ggataagagg ctctggggat tgacgtgagg gggcagggat 60  
 ggctatattt ctgggagcga actccgggcg a 91

<210> 47  
 <211> 71  
 <212> DNA  
 <213> Nicotiana tabacum

<220>  
 <221> promoter  
 <222> (1)..(71)  
 <223> plastidic promoter PaccD-129

<400> 47  
 gtcgacatat tattttaaataaatataaagg ggggtccaac atattaatat atagtgaagt 60  
 gttccggatc c 71

<210> 48  
 <211> 27  
 <212> DNA  
 <213> Nicotiana tabacum

<220>  
 <221> promoter  
 <222> (1)..(27)  
 <223> plastidic promoter PclpP-53

<400> 48  
 agacaataaa aaaaattggt acgtttc 27

<210> 49  
 <211> 65  
 <212> DNA  
 <213> Nicotiana tabacum

<220>  
 <221> promoter  
 <222> (1)..(65)  
 <223> plastidic promoter Prn-62

<400> 49  
 gagcgaactc cgggcgaata tgaagcgcat ggatacaagt tatgccttgg aatgaaagac 60  
 aattc 65

<210> 50  
 <211> 107  
 <212> DNA  
 <213> Nicotiana tabacum

<220>  
 <221> promoter  
 <222> (1)..(107)  
 <223> plastidic promoter Prps16

0817/00028

23

&lt;400&gt; 50

tctatata tttatatatt aatatattaat taatatgtta gataaattaa tattagcgat 60  
 ggggtcttac taaagaaaaa tatttatcca cctatctcta tagtata 107

&lt;210&gt; 51

&lt;211&gt; 47

&lt;212&gt; DNA

&lt;213&gt; Nicotiana tabacum

&lt;220&gt;

&lt;221&gt; promoter

&lt;222&gt; (1)..(47)

&lt;223&gt; plastidic promoter PatpB/E-290

&lt;400&gt; 51

agaaatagaa aataaagttc aggttcgaat tccatagaat agataat 47

&lt;210&gt; 52

&lt;211&gt; 25

&lt;212&gt; DNA

&lt;213&gt; Nicotiana tabacum

&lt;220&gt;

&lt;221&gt; promoter

&lt;222&gt; (1)..(25)

&lt;223&gt; plastidic promoter PrpoB-345

&lt;400&gt; 52

aatgtgtatt atcataataa tggta 25

&lt;210&gt; 53

&lt;211&gt; 35

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence: consensus  
 sequence of E.coli sigma70 promoter

&lt;400&gt; 53

ttgacattca ctctcaatt atctataatg ataca 35

&lt;210&gt; 54

&lt;211&gt; 80

&lt;212&gt; DNA

&lt;213&gt; Nicotiana tabacum

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1)..(80)

<223> 5'-untranslated region of psbA gene (incl. NcoI  
 site)

0817/00028

24

<400> 54  
 tccattttct atttgattt gtagaaaact agtgtgcttg ggagtcctg atgattaaat 60  
 aaaccaagat ttaccatgg 80

<210> 55  
 <211> 40  
 <212> DNA  
 <213> Nicotiana tabacum  
 <220>  
 <221> misc\_feature  
 <222> (1)..(40)  
 <223> 5'-untranslated region of rbcL gene

<400> 55  
 agttgtaggg agggattcat gacaccacaa acagccatgg 40

<210> 56  
 <211> 22  
 <212> DNA  
 <213> Nicotiana tabacum  
 <220>  
 <221> misc\_feature  
 <222> (1)..(22)  
 <223> 5'-untranslated region of rbcLs gene (modified)

<400> 56  
 agttgtaggg agggattcat ga 22

<210> 57  
 <211> 127  
 <212> DNA  
 <213> Synechocystis sp.  
 <220>  
 <221> misc\_feature  
 <222> (1)..(127)  
 <223> 3'-untranslated region of synechcystis psbA-1 gene

<400> 57  
 tgccattgcc ataactgctt tcggttagac ttcgtttcat ttggttaac aagggcactc 60  
 tcgcaatggg gtgccttta tggccaagg ttaaagttaa gccagtacta ttctagggt 120  
 gaaatgt 127

<210> 58  
 <211> 90  
 <212> DNA  
 <213> Nicotiana tabacum  
 <220>  
 <221> misc\_feature  
 <222> (1)..(90)  
 <223> 3'-untranslated region of tobacco psbA gene

0817/00028

25

&lt;400&gt; 58

cctggcctag tctataggag gtttgaaaa gaaaggagca ataatcattt tctgttcta 60  
 tcaagagggt gctattgctc ctttctttt 90

&lt;210&gt; 59

&lt;211&gt; 136

&lt;212&gt; DNA

&lt;213&gt; Nicotiana tabacum

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1)..(136)

&lt;223&gt; 3'-untranslated region of tobacco rbcL gene

&lt;400&gt; 59

agtagacatt agcagataaa ttagcaggaa ataaagaagg ataaggagaa agaactcaag 60  
 taatatcct tcgttctctt aattgaattg caattaaact cggcccaatc tttaactaaa 120  
 aggattgagc cgaata 136

&lt;210&gt; 60

&lt;211&gt; 13

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence: synthetic  
 ribosome binding site (RBS)

&lt;400&gt; 60

ggaggnnnnn atg 13

&lt;210&gt; 61

&lt;211&gt; 13

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence: synthetic  
 ribosome binding site (RBS)

&lt;400&gt; 61

ggaggatctc atg 13

&lt;210&gt; 62

&lt;211&gt; 1494

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence: complete  
 insert of vector pCB304-25

&lt;400&gt; 62

ggtaccgggc cgataattcc tggcctagtc tataggaggt ttgaaaaga aaggagcaat 60

0817/00028

26

aatcatttc tigtctalc aagagggtgc tattgctcct ttctttttt ctcgagtacg 120  
 ctagggataa cagggtataa tagaagcttg actctcttaa agatctcgac tctatatata 180  
 ttatatatt aaatattaat taatatgta gataaattaa tattagcgat ggggtcttac 240  
 taaagaaaaa tatttatcca cctatcicta tagtatatag gatctagttag tagggaggga 300  
 ttcatgacac cacaacagc catgggagct tggatigaac aagatggalt gcacgcaggt 360  
 tctcgggccg cttgggtgga gaggtattc ggctatgact gggcacaaca gacaatcggc 420  
 tgctctgatg ccgccgtgtt ccggctgtca gcgcaggggc gcccggttct tttgtcaag 480  
 accgacctgt ccggtgocct gaatgaactg caggacgagg cagcgcggtc atcgtggctg 540  
 gccacgacgg gcgttccttg cgcagctgtg ctcgacgttg tcactgaagc gggaaggagc 600  
 tggctgctat tgggcgaagt gccggggcag gatctcctgt catctcacct tgctcctgcc 660  
 gagaagtat ccatcatggc tgaatgaatg cggcggtgtc atacgcttga tccggctacc 720  
 tgccattcgc accaccaagc gaaacatcgc atcgagcgag cacgtactcg gatggaagcc 780  
 ggctctgtc atcaggatga tctggacgaa gagcatcagg ggctcgcgcc agccgaactg 840  
 ttcgccagc tcaaggcgcg catgcccgac ggcgaggatc tcgtcgtgac ccatggcgat 900  
 gcctgcttgc cgaatatcat ggtggaaaat ggccgctttt ctggattcat cgactgtggc 960  
 cggctgggtg tggcggaacc ctatcaggac atagcgttgg ctaccctga tattgctgaa 1020  
 gagcttggcg gcgaatgggc tgaccgcttc ctcgtgcttt acggtatcgc cgctcccgat 1080  
 tcgcagcgca tcgccttcta tcgccttctt gacgagtctt tctgagcggg acccaagcta 1140  
 ggcgcccgcg gatcaagct tgaattcagt agacattagc agataaatta gcaggaaata 1200  
 aagaaggata aggagaaaga actcaagtaa ttatccttcg ttctcttaat tgaattgcaa 1260  
 ttaaactcgg cccaatctt tactaaaagg attagccga atactcgact taaggtagcg 1320  
 tcgagtacgc gcccggggag cccaagggca gcgcctggca cccgaagctc tagtatcaa 1380  
 ttggcaca aaagcaaat taaatactg ataattgcca acacaattaa catctcaatc 1440  
 aaggtaaatg cttttgctt ttttgccaa agctatcttc cgtgatcaga gctc 1494

&lt;210&gt; 63

&lt;211&gt; 592

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence: BglII / MnlI  
 fragment from vector pCB320-192

&lt;400&gt; 63

agatcttcca ttcttattt tgattgtag aaaactagt tgctggggag tccctgatga 60  
 ttaaataaac caagatttta ccatggcgct caccaatgct caaatcttgg ctgtgattga 120  
 cagttgggaa gaaacagtcg gtcagtttcc agtgataacg caccatgtac cattaggtgg 180  
 cggctcga gaaacgctcc attgttacga gatccoccta gcagctcctt atgggggttg 240  
 ctttgctaag aatgggccta cccgctggca atacaaacgg acaatcaatc aagtcgtcca 300  
 cagatggggg tccacacag tccctttct attagaaccg gataacatca acggcaaac 360  
 ctgcacagca tcgcacctat gtcataatac tcgatgocac aatcccttgc acttgctg 420  
 ggagtcacta gacgacaaca aaggcagaaa ctgggtgccc ggtoccaaag ggggatgtgt 480  
 ccatgcggtg gttgtttaa ggcagggtcc gttgtacggc ccaggggcga ctgtggcagg 540  
 tctcaaaa aggggcagtc acttgtggt ataactgcag aagcttcaat tg 592

&lt;210&gt; 64

&lt;211&gt; 25

0817/00028

27

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence:  
oligonucleotide primer

&lt;400&gt; 64

aaagatctcc tcacaaaggg ggtcg

25

&lt;210&gt; 65

&lt;211&gt; 23

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence:  
oligonucleotide primer

&lt;400&gt; 65

tcgaagactt aggaccgtta tag

23

&lt;210&gt; 66

&lt;211&gt; 26

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence:  
oligonucleotide primer

&lt;400&gt; 66

aggaagacct tgcgggtaa gttccg

26

&lt;210&gt; 67

&lt;211&gt; 31

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence:  
oligonucleotide primer

&lt;400&gt; 67

ctcaattggg gtctctgt ccaggtgcag g

31

&lt;210&gt; 68

&lt;211&gt; 672

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence: artificial  
fusion protein of I-Ppo I homing endonuclease and  
plastidic transit peptide



0817/00028

28

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)..(669)

&lt;400&gt; 68

atg tct gct tct tct tta ttt aat ctc cca ttg att cgc ctc aga tct 48

Met Ser Ala Ser Ser Leu Phe Asn Leu Pro Leu Ile Arg Leu Arg Ser

1 5 10 15

ctc gct ctt tgg tct tct ttt tct ttc cga ttt gcc cat cgt cct 96

Leu Ala Leu Ser Ser Ser Phe Ser Ser Phe Arg Phe Ala His Arg Pro

20 25 30

ctg tca tgg att tca ccg aga aag tta ccg aat ttt cgt gct ttc tct 144

Leu Ser Ile Ser Pro Arg Lys Leu Pro Asn Phe Arg Ala Phe Ser

35 40 45

ggg acc gct atg aca gat act aaa gat ggg atc ccc atg gcg ctc acc 192

Gly Thr Ala Met Thr Asp Thr Lys Asp Gly Ile Pro Met Ala Leu Thr

50 55 60

aat gct caa atc ttg gct gtg att gac agt tgg gaa gaa aca gtc ggt 240

Asn Ala Gln Ile Leu Ala Val Ile Asp Ser Trp Glu Glu Thr Val Gly

65 70 75 80

cag ttt cca gtg ata acg cac cat gta cca tta ggt ggc ggt ctg caa 288

Gln Phe Pro Val Ile Thr His His Val Pro Leu Gly Gly Gly Leu Gln

85 90 95

gga acg ctc cat tgt tac gag atc ccc cta gca gct cct tat ggg gtt 336

Gly Thr Leu His Cys Tyr Glu Ile Pro Leu Ala Ala Pro Tyr Gly Val

100 105 110

ggc ttt gct aag aat ggg cct acc cgc tgg caa tac aaa cgg aca atc 384

Gly Phe Ala Lys Asn Gly Pro Thr Arg Trp Gln Tyr Lys Arg Thr Ile

115 120 125

aat caa gtc gtc cac aga tgg ggg tcc cac aca gtc cct ttt cta tta 432

Asn Gln Val Val His Arg Trp Gly Ser His Thr Val Pro Phe Leu Leu

130 135 140

gaa ccg gat aac atc aac ggc aaa acc tgc aca gca tgg cac cta tgt 480

Glu Pro Asp Asn Ile Asn Gly Lys Thr Cys Thr Ala Ser His Leu Cys

145 150 155 160

cat aat act cga tgc cac aat ccc ttg cac ttg tgc tgg gag tca cta 528

His Asn Thr Arg Cys His Asn Pro Leu His Leu Cys Trp Glu Ser Leu

165 170 175

gac gac aac aaa ggc aga aac tgg tgc ccc ggt ccc aac ggg gga tgt 576

Asp Asp Asn Lys Gly Arg Asn Trp Cys Pro Gly Pro Asn Gly Gly Cys

180 185 190

gtc cat gcg gtg gtt tgt tta agg cag ggt ccg ttg tac ggc cca ggg 624

Val His Ala Val Val Cys Leu Arg Gln Gly Pro Leu Tyr Gly Pro Gly

195 200 205

0817/00028

29

gcg act gtg gca ggt cct caa caa agg ggc agt cac ttt gtg gta taa 672  
 Ala Thr Val Ala Gly Pro Gln Gln Arg Gly Ser His Phe Val Val  
 210 215 220

&lt;210&gt; 69

&lt;211&gt; 223

&lt;212&gt; PRT

&lt;213&gt; Artificial sequence

<223> Description of the artificial sequence: artificial  
 fusion protein of I-Ppo I homing endonuclease and  
 plastidic transit peptide

&lt;400&gt; 69

Met Ser Ala Ser Ser Leu Phe Asn Leu Pro Leu Ile Arg Leu Arg Ser  
 1 5 10 15

Leu Ala Leu Ser Ser Ser Phe Ser Ser Phe Arg Phe Ala His Arg Pro  
 20 25 30

Leu Ser Ser Ile Ser Pro Arg Lys Leu Pro Asn Phe Arg Ala Phe Ser  
 35 40 45

Gly Thr Ala Met Thr Asp Thr Lys Asp Gly Ile Pro Met Ala Leu Thr  
 50 55 60

Asn Ala Gln Ile Leu Ala Val Ile Asp Ser Trp Glu Glu Thr Val Gly  
 65 70 75 80

Gln Phe Pro Val Ile Thr His His Val Pro Leu Gly Gly Gly Leu Gln  
 85 90 95

Gly Thr Leu His Cys Tyr Glu Ile Pro Leu Ala Ala Pro Tyr Gly Val  
 100 105 110

Gly Phe Ala Lys Asn Gly Pro Thr Arg Trp Gln Tyr Lys Arg Thr Ile  
 115 120 125

Asn Gln Val Val His Arg Trp Gly Ser His Thr Val Pro Phe Leu Leu  
 130 135 140

Glu Pro Asp Asn Ile Asn Gly Lys Thr Cys Thr Ala Ser His Leu Cys  
 145 150 155 160

His Asn Thr Arg Cys His Asn Pro Leu His Leu Cys Trp Glu Ser Leu  
 165 170 175

Asp Asp Asn Lys Gly Arg Asn Trp Cys Pro Gly Pro Asn Gly Gly Cys  
 180 185 190

Val His Ala Val Val Cys Leu Arg Gln Gly Pro Leu Tyr Gly Pro Gly  
 195 200 205

Ala Thr Val Ala Gly Pro Gln Gln Arg Gly Ser His Phe Val Val  
 210 215 220

0817/00028

30

&lt;210&gt; 70

&lt;211&gt; 558

&lt;212&gt; DNA

&lt;213&gt; Physarum polycephalum

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)..(555)

&lt;223&gt; coding for long version of I-PpoI

&lt;400&gt; 70

atg gcg aaa tcc aac caa gct cgg gta aac ggc ggg agt aac tat gac 48  
Met Ala Lys Ser Asn Gln Ala Arg Val Asn Gly Gly Ser Asn Tyr Asp

1 5 10 15

tct ctc acc ccc tta aat atg gcg ctc acc aat gct caa atc ttg gct 96  
Ser Leu Thr Pro Leu Asn Met Ala Leu Thr Asn Ala Gln Ile Leu Ala

20 25 30

gtg att gac agc tgg gaa gaa aca gtc ggt cag ttt cca gtg ata acg 144  
Val Ile Asp Ser Trp Glu Glu Thr Val Gly Gln Phe Pro Val Ile Thr

35 40 45

cac cat gta cca tta ggt ggc ggt ctg caa gga acg ctc cat tgt tac 192  
His His Val Pro Leu Gly Gly Gly Leu Gln Gly Thr Leu His Cys Tyr

50 55 60

gag atc ccc cta gca gct cct tat ggg gtt ggc ttt gct aag aat ggg 240  
Glu Ile Pro Leu Ala Ala Pro Tyr Gly Val Gly Phe Ala Lys Asn Gly

65 70 75 80

cct acc cgc tgg caa tac aaa cgg aca atc aat caa gtc gtc cac aga 288  
Pro Thr Arg Trp Gln Tyr Lys Arg Thr Ile Asn Gln Val Val His Arg

85 90 95

tgg gga tcc cac aca gtc cct ttt cta tta gaa ccg gat aac atc aac 336  
Trp Gly Ser His Thr Val Pro Phe Leu Leu Glu Pro Asp Asn Ile Asn

100 105 110

ggc aaa acc tgc aca gca tgc cac cta tgt cat aat act cga tgc cac 384  
Gly Lys Thr Cys Thr Ala Ser His Leu Cys His Asn Thr Arg Cys His

115 120 125

aat ccc ttg cac ttg tgc tgg gag tca cta gac gac aac aaa ggc aga 432  
Asn Pro Leu His Leu Cys Trp Glu Ser Leu Asp Asp Asn Lys Gly Arg

130 135 140

aac tgg tgc ccg ggt ccc aac ggg gga tgt gtc cat gcg gtg gtt tgt 480  
Asn Trp Cys Pro Gly Pro Asn Gly Gly Cys Val His Ala Val Val Cys

145 150 155 160

tta agg cag ggt ccg ttg tac ggc ccg ggg gcg act gtg gca ggt cct 528  
Leu Arg Gln Gly Pro Leu Tyr Gly Pro Gly Ala Thr Val Ala Gly Pro

165 170 175

0817/00028

31

caa caa agg ggc agt cac ttt gtg gta taa 558  
 Gln Gln Arg Gly Ser His Phe Val Val  
 180 185

&lt;210&gt; 71

&lt;211&gt; 185

&lt;212&gt; PRT

&lt;213&gt; Physarum polycephalum

&lt;400&gt; 71

Met Ala Lys Ser Asn Gln Ala Arg Val Asn Gly Gly Ser Asn Tyr Asp  
 1 5 10 15

Ser Leu Thr Pro Leu Asn Met Ala Leu Thr Asn Ala Gln Ile Leu Ala  
 20 25 30

Val Ile Asp Ser Trp Glu Glu Thr Val Gly Gln Phe Pro Val Ile Thr  
 35 40 45

His His Val Pro Leu Gly Gly Gly Leu Gln Gly Thr Leu His Cys Tyr  
 50 55 60

Glu Ile Pro Leu Ala Ala Pro Tyr Gly Val Gly Phe Ala Lys Asn Gly  
 65 70 75 80

Pro Thr Arg Trp Gln Tyr Lys Arg Thr Ile Asn Gln Val Val His Arg  
 85 90 95

Trp Gly Ser His Thr Val Pro Phe Leu Leu Glu Pro Asp Asn Ile Asn  
 100 105 110

Gly Lys Thr Cys Thr Ala Ser His Leu Cys His Asn Thr Arg Cys His  
 115 120 125

Asn Pro Leu His Leu Cys Trp Glu Ser Leu Asp Asp Asn Lys Gly Arg  
 130 135 140

Asn Trp Cys Pro Gly Pro Asn Gly Gly Cys Val His Ala Val Val Cys  
 145 150 155 160

Leu Arg Gln Gly Pro Leu Tyr Gly Pro Gly Ala Thr Val Ala Gly Pro  
 165 170 175

Gln Gln Arg Gly Ser His Phe Val Val  
 180 185

&lt;210&gt; 72

&lt;211&gt; 34

&lt;212&gt; DNA

&lt;213&gt; Escherichia coli

&lt;220&gt;

&lt;221&gt; promoter

&lt;222&gt; (1)..(34)

<223> coding for promoter derived from consensus of  
 sigma70 E.coli promoters

0817/00028

32

&lt;400&gt; 72

ttgacaattt tctctgaat tatataatta acat

34

&lt;210&gt; 73

&lt;211&gt; 413

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence: artificial  
intron TetIVS2a

&lt;400&gt; 73

aaatagcaat attttctgtt gccttaaaaa gttatcaggc atgcacctgg tagctagtct 60  
 ttaaaccaat agattgcacg ggtttaaaag gcaagaccgt caaattgcgg gaaaggggtc 120  
 aacagccgtt cagtaccaag tctcagggga aactttgaga tggccttgca aagggtatgg 180  
 taataagctg acggacatgg tcttaaccac gcagccaagt cctaagtcaa cagatcttct 240  
 gttgatatgg atgcagttca cagactaaat gtcggtcggg gaagatgtat tcttctcata 300  
 agatatagtc ggacctctcc ttaatgggag ctagcggatg aagtgatgca acactggagc 360  
 cgtcgggaac taattgtat gcgaaagtat attgattagt ttggagtagc tcg 413

&lt;210&gt; 74

&lt;211&gt; 791

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence: insert of  
plasmid pCB459-1

&lt;400&gt; 74

ttacaatttc cattcgccat tcaggctgcg caactgttgg gaagggcgat cggtcggggc 60  
 ctcttcgcta ttacgccagc tggcgaaagg gggatgtgct gcaaggcgat taagtgggt 120  
 aacgccaggg tttccagc cagcagcttg taaaacgacg gccagtgagc gcgcgtaata 180  
 cgactcacta tagggcgaat tgggtaccgg gccccccctc gaggtcgacg gtatcgataa 240  
 gcttgaccgg acaaggaatt tcgctcgagt actccaaaac taatcaatat acttgcgat 300  
 acaaattagt tccagcggc tccagtgtg catcacttca tccgctagct cccattaagg 360  
 agaggctcga ctatatctta tgagaagaat acatcttccc cgaccgacat ttagtctgtg 420  
 aactgcatcc atatcaacag aagatcigt gacttaggac ttggctgcgt ggtaggacc 480  
 atgtccgtca gcttattacc atacccttg caaggccatc tcaaagtttc ccctgagact 540  
 tggtagtgaa cggctgttga ccccttccc gcaattgac ggtctgcct tttaaaccga 600  
 tgcaatctat tggtttaaag actagctacc aggtgcatgc ctgataactt ttaaggcaa 660  
 cgaaaaatat tgctatttac cttaggaccg ttaggatcca ctagtctag agcggccgcc 720  
 accgcggtgg agctccagct ttgttcctt ttagtgaggg ttaattgcgc gcttggcgta 780  
 atcatggtca t 791

&lt;210&gt; 75

&lt;211&gt; 1291

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

0817/00028

33

&lt;220&gt;

<223> Description of the artificial sequence: insert of  
vector pCB478-3

&lt;400&gt; 75

ttacaatttc cattcgccat tcaggctgcg caactgttgg gaagggcgat cggcgcgggc 60  
 ctcttcgcta ttacgccagc tggcgaaagg gggatgtgct gcaaggcgat taagtgggt 120  
 aacgccaggg ttitcccagt cagcacgttg taaaacgacg gccagtgagc gcgcgtaata 180  
 cgactcacta tagggcgaat tgggtaccgg gccccccctc gaggtcgacg gtatcgataa 240  
 gcttgaccog acaaggaatt tcgctcgagt actccaaaac taatcaatat actttcgcat 300  
 acaaattagt tcccagcggc tccagtgttg catcacttca tccgctagct cccattaagg 360  
 agagggtccga ctatatctta tgagaagaat ttgatacctt actcgccagt ttcttcaaaa 420  
 cgtttgatag gtctctgcta gtttttctca ttgatacgga gttttttgc taaatccaac 480  
 actttctcga atccttcctg cgattctaaa tgtctacctg cttttaacag caagcagatt 540  
 ttcgaaatc gaagaaatc aattctctt ttgttttta gctcatgtt ttcgaaaaat 600  
 ggtataatga cgggtataaa gtgatcaata ttttgactt tatagcaaag ccgttcccca 660  
 tgattcaca cgactgatcc aactttaag tagtcttta atgcatataa aacatttga 720  
 tctaatttat gttgggtcac tacgaattaa ttctggttga agctgatggc catagcgcaa 780  
 cgaattattt ctaagtatac ttactataa acaaccttcc ccactctacg aacctgtaat 840  
 ccattgagga ttaatgtcca tggtcctcc ctacaacgct tgcagatcaa acatcttccc 900  
 cgaccgacat ttagtctgtg aactgcatcc atatcaacag aagatctgtt gacttaggac 960  
 ttggctgctg ggttaggacc atgtccgtca gcttattacc atacctttg caaggccatc 1020  
 tcaaagtttc cctgagact tggactgaa cggctgttga ccccttccc gcaatttgac 1080  
 ggtcttgctt tttaaacga tgcaatctat tggtttaaag actagctacc aggtgcatgc 1140  
 ctgataactt tttaaggcaa cgaaaaatat tgctatttac cttaggaccg ttaggatcca 1200  
 ctagtctag agcggccgcc accgcggtgg agctccagct ttgttccct ttagtgaggg 1260  
 ttaattgcgc gcttggcgta atcatggta t 1291

&lt;210&gt; 76

&lt;211&gt; 1386

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence: insert of  
vector pCB492-25

&lt;400&gt; 76

gcgcgcgtaa tacgactcac tatagggcga attgggggcg actgtttacc aaaaacacag 60  
 gtctccgcaa agtcgtaaga ccatgtatgg gggctgacgc ctgccagtg ccggaaggct 120  
 aagggaagtg gtgacctgat gacaggggag cggcgaccg aagccccggt gaacggcggc 180  
 cgttaactata acggtcctaa ggtaaatagc aatattttc gttgccttaa aaagttatca 240  
 ggcatgcacc tggtagctag tctttaaac aatagattgc atcggtttaa aaggcaagac 300  
 cgtcaaattg cgggaaagg gtaacagcc gttcagtacc aagtctcagg ggaactttg 360  
 agatggcctt gcaaagggtg tggtaataag ctgacggaca tggctctaac cagcagcca 420  
 agtctaagt caacagatct tctgttgata tggatgcagt tcacagacta aatgtcggtc 480  
 ggggaagatg ttgatctgc agacgttgta gggagggacc atggacatta atcctcaatg 540  
 gattacaggt ttcgtagatg gggaagggtg ttttagtgta agtatactta gaaataatc 600  
 gttagcctat ggcatcagc ttcaaccaga attcgtagt acccaacata aattagatgc 660

0817/00028

34

aaatgtttta tatgcattaa aagactactt taaagttgga tcagtcgttg tgaatcatgg 720  
 ggaacgggctt tgctataaag tcaaaaatat tgatcacttt ataaccgtca ttataccatt 780  
 tticgaaaaa catgagctaa aaacaaaaag aagaattgaa ttcttcgat ttgaaaaat 840  
 ctgcttgctg ttaaaagcag gtagacattt agaalcgcag gaaggattcg agaaagtgtt 900  
 ggatttagca aaaaaactcc gtatcaatga gaaaaactac caggaatcta tcaaacgttt 960  
 tgaagaaact ggcgagtaag gatcaaaatt cttctcataa gatatagtcg gacctctcct 1020  
 taatgggagc tagcggatga agtgatgcaa cactggagcc gctgggaact aatttgtatg 1080  
 cgaaagtata ttgattagtt ttggagtact cgagcgaaat tccttgctcg gtaagttccg 1140  
 acccgcacga aaggcgtaac gatctgggca ctgtctcgga gagaggctcg gtgaaataga 1200  
 catgtctgtg aagatgcgga ctacctgcac ctggacagag agacccaat tcttggccta 1260  
 gtctatagga ggtttgaaa agaaaggagc aataatcatt ttctgttct atcaagaggg 1320  
 tgctattgct cctttcttt ttctcgacc agctttgtt cccttagtg agggtaatt 1380  
 gcgcgc 1386

&lt;210&gt; 77

&lt;211&gt; 60

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

&lt;223&gt; Description of the artificial sequence:

oligonucleotide primer

&lt;400&gt; 77

tagtaaatga caattttcct ctgaattata taattaacat ggcgactgtt taccaaaaac 60

&lt;210&gt; 78

&lt;211&gt; 31

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

&lt;223&gt; Description of the artificial sequence:

oligonucleotide primer

&lt;400&gt; 78

ctcaattggg gtctctgt ccagggtcag g 31

&lt;210&gt; 79

&lt;211&gt; 1258

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

&lt;223&gt; Description of the artificial sequence: PCR product

Prom-TetIVS2a-Cpa

&lt;400&gt; 79

tagtaaatga caattttcct ctgaattata taattaacat ggcgactgtt taccaaaaac 60  
 acagggtctcc gcaaaagtcgt aagaccatgt atgggggctg acgcctgccc agtgccggaa 120  
 ggtaagga gttggtgacc tgatgacagg ggagccggcg accgaagccc cggtaaacgg 180  
 cggccgtaac tataacggtc ctaaggtaaa tagcaatatt ttctgttgc ttaaaaagtt 240

0817/00028

35

atcaggcatg cacctggtag ctagtcttta aaccaataga tgcacggt ttaaaaggca 300  
 agaccgtcaa attgcgggaa aggggtcaac agccgttcag taccaagtct caggggaaac 360  
 ttgagatgg ccttgcaaag ggtatggtaa taagctgacg gacatggloc taaccacgca 420  
 gccaaagtct aagtcaacag atcttctgtt gatatggatg cagttcacag actaaatgtc 480  
 ggtcggggaa gatgtttgat ctgcagacgt ttagggagg gaccatggac attaatcttc 540  
 aatggattac aggtttcgtg gatggggaag gttgttttag tgaagtata cttagaaata 600  
 attcgttgcg ctatggccat cagcttcaac cagaattcgt agtgacccaa cataaattag 660  
 atgcaaatgt ttatatgca ttaaaagact actttaaagt tggatcagtc gttgtgaatc 720  
 atggggaacg gctttgctat aaagtcaaaa atattgatca ctttataacc gtcattatc 780  
 cattttgcg aaaacatgag ctaaaaacaa aaagaagaat tgaatttctt cgatttcgaa 840  
 aaatctgctt gctgttaaaa gcaggtagac atttagaatc gcaggaagga ttcgagaaag 900  
 tgttggaatt agcaaaaaaa ctccgtatca atgagaaaaa ctaccaggaa tctatcaaac 960  
 gttttgaaga aactggcgag taaggatcaa aattcttctc ataagataa gtcggacctc 1020  
 tcttaatgg gagctagcgg atgaagtgat gcaacactgg agccgtcggg aactaattg 1080  
 tatgcgaaa tatattgatt agtttgag tactcgagcg aaattccttg tcgggtaagt 1140  
 tccgaccgc acgaaaggcg taacgatctg ggactgtct cggagagagg ctccgtgaaa 1200  
 tagacatgtc tgtgaagatg cggactacct gcacctggac agagagacc caattgag 1258

&lt;210&gt; 80

&lt;211&gt; 4746

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence: insert of  
 vector pCB435-45

&lt;400&gt; 80

gcgccaata cgaaaccgc ctctcccgcg gcgttggcgg attcattaat gcagtatga 60  
 ctgacggca tctcacctt cctccggctt atcaccggca gtcgttcag ggttccaaac 120  
 tcaacgatgg caactaaaca cgagggttgc gtcgttgcg ggacttaacc caacacctta 180  
 cggcacgagc tgacgacagc catgcaccac ctgtgtccgc gttccgaag gcacctctc 240  
 ctticaagag gattcggcg atgtcaagcc ctggttaagg tcttcgctt gcatcgaatt 300  
 aaaccacatg ctccaccgct tgtcggggcc cccgtcaatt ccttttagt tcatcttgc 360  
 gaacgtactc cccaggcggg atacttaacg cgttagctac agcactgcac gggtcgatac 420  
 gcacagcgcc tagtatccat cgtttacggc taggactact ggggtatcta atccattcg 480  
 ctccccagc ttctgtctc cagtgctagt gtcggccag cagagtgtt tcgccgttg 540  
 tgttcttcc gatctctac catttcaccg ctccaccgga aattccctc gccctaccg 600  
 tactccagct tggtagttc caccgcctgt ccagggttga gccctgggat ttgacggcgg 660  
 acttaaaaag ccacctacag acgctttacg ccaatcatt ccggataacg ctgcatcct 720  
 ctgtattacc cgggctgctg gcacagagt agccgatgt tattccccag ataccgtcat 780  
 tcttctctc ccgggaaaag aagttcacga cccgtgggccc ttctacctc acgcggcatt 840  
 gtcocgtcag gcttgcgcc attgcggaaa attccccact gtcgctccc gtaggagtct 900  
 gggccgtgtc tcagtccag tgtggctgat catctctcg gaccagctac tgatcatcg 960  
 ctggttaagc tattgcctca ccaactagct aatcagacgc gagccctcc tcgggcggat 1020  
 tctctctt gctctcagc ctacggggtg ttagcagccg ttccagctg ttgtccct 1080  
 cccaagggca ggttcttacg cgttactcac ccgtccgcca ctggaacac cactccct 1140  
 ccgacttga tgtgtaagc atccgccag cgttcatct gagccaggat cgaactctc 1200  
 atgagatca tagttgcat acttatagct tcttgttcg tagacaaagc ggattcgaa 1260



0817/00028

36

ttgtcttca ttccaaggca taacttgtat ccatgcgctt catattgcc cggagttcgc 1320  
 tcccagaaat atagccatcc ctgccccctc acgtcaatcc cagcagcctc ttatccattc 1380  
 tcattgaacg acggcggggg agcgcggggc ccccccgac atatatttta aataatataa 1440  
 aggggggttc aacatattaa tatatagtga agtgtccgg atccactagg acgttgtagg 1500  
 gagggaccat ggacattaat cctcaatgga ttacagggtt cgtagatggg gaagggtgtt 1560  
 ttagtgaag tatacttaga aataatcgt tgcgctatgg ccatcagctt caaccagaat 1620  
 tcgtagtgc ccaacataaa ttgatgcaa atgttttata tgcattaaa gactacttta 1680  
 aagttggatc agtcgtgtg aatcatgggg aacggcttg ctataaagtc aaaaatattg 1740  
 atcactttat aaccgtcatt ataccattt tcaaaaaa tgagctaaaa acaaaaagaa 1800  
 gaattgaatt tcttcgattt cgaaaaatct gcttgctgtt aaaagcaggt agacatttag 1860  
 aatcgcagga aggattcgag aaagtgttg atttagcaaa aaaactccgt atcaatgaga 1920  
 aaaactacca ggaatctatc aaacgtttg aagaactgg cgagtaactc gagaaaaaaa 1980  
 gaaaggagca atagcacctt ctgatagaa caagaaaatg attattgctc ctttctttc 2040  
 aaaactctt atagactagg ccaggaatta tctgcagtta ttgccaact acctagtaga 2100  
 tctgccttt cactagtagg acaattctt ccaactgac tgcgcgcgag gccaaagcat 2160  
 cttctcttg tcaagataa gcctgtctag ctcaagat gacgggctga tactgggccc 2220  
 gcaggcgctc cattgcccag tggcagcga catcctcgg cgcgatttg ccggttactg 2280  
 cgctgtacca aatgcgggac aacgtaagca ctacatttc ctcacgcca gccagtcgg 2340  
 gcggcgagtt ccatagcgtt aaggttcat ttagegcctc aaatagatcc tgttcaggaa 2400  
 ccggatcaaa gagttcctcc gccgtggac ctaccaaggc aacgctatgt tctctgtt 2460  
 ttgtcagcaa gatagccaga tcaatgtcga tctggctgg ctgaagata cctgcaagaa 2520  
 tgtcatgctg ctgccattct ccaaattgca gttcgcgctt agctggataa cgccacggaa 2580  
 tgatgtcgtc gtgcacaaca atggtgactt ctacagcgcg gagaatctcg ctctctcag 2640  
 gggaagccga agtttcaaa aggtcgttga tcaaagctcg ccgctgtgtt tcatcaagcc 2700  
 ttacggcac cgtaacccagc aaatcaatat cactgtgtgg ctacaggccg ccatccactg 2760  
 cggagccgta caaatgtacg gccagcaacg tgggttcgag atggcgctcg atgacgcaa 2820  
 ctactctga tagttgagtt gatacttcgg cgataaccgc ttacagagcc atgaatccct 2880  
 cctacaact agatcctcgc ccggagttcg ctcccagaaa tatagccatc cctgccccct 2940  
 cactgaatc ccacgagcct ctatccatt ctcatgaac gacggcgctc aggggggggc 3000  
 cggtaacaaa tccaactaga aaaactcaca ttgggttag ggataatcag gctcgaactg 3060  
 atgacttcca ccacgtcaag gtgacactct accgtgagt tatatccctt ccccgcccca 3120  
 tcgagaaata gaactgacta atcctaagtc aaagggtcga gaaactcaac gccactattc 3180  
 ttgaacaact tggagccggg cttcttttc gcactattac ggatagaaa ataattgtca 3240  
 aaatcggtt caattgtcaa ctgcccctat cggaaatagg attgactacc gattccgaag 3300  
 gaactggagt tacatctctt ttcatcaca gatttctat gcgtttccac gccctttga 3360  
 gaccccgaaa aatggacaaa ttcttttct taggaacaca tacaagattc gtcactaca 3420  
 aaaggataat ggtaacccta ccattaacta ctcatattt gaatttcata gtaatagaaa 3480  
 tacatgtcct accgagacag aatttgaac ttgctatcct ctgctagc aggcaaagat 3540  
 ttacctcgt ggaaaggatg attcattcgg atcgacatga gagtccaact acattgccag 3600  
 aatccatgt gtatattga aagaggttga cctccttgc tctctcatgg tacactcctc 3660  
 ttcccgcca gcccttttc tctcgggtcc acagagacaa aatgtaggac tgggtccaac 3720  
 aattcatcag actcactaag tgggatcac taactaata taatctaata taatagtcta 3780  
 atatactaa tataatagaa aatactaata taatagaaaa gaactgtctt ttctgtatac 3840  
 ttccccgt tccgttgcta ccgagggcct cgtgatacgc ctattttat aggttaatgt 3900  
 catgacaaa atccctaac gtgagtttc gtccactga gcgtcagacc ccgtagaaaa 3960  
 gatcaaagga tcttcttgag atccttttt tctgcgcgta atctgtctgt tgcaacaaaa 4020

0817/00028

37

aaaaccaccg ctaccagcgg tggttgttt gccggatcaa gagctaccaa ctcctttcc 4080  
 gaaggttaact ggcttcagca gagcgcagat accaaatact gtccttctag ttagccgta 4140  
 gtaggccac cacttcaaga actctgtagc accgcctaca tacctcgtc tgctaactct 4200  
 gttaccagtg gctgctgcca gtggcgataa gtcgtgtctt accgggttgg actcaagacg 4260  
 atagtiaccg gataaggcgc agcggtcggg ctgaacgggg ggttcgtgca cacagcccag 4320  
 ctgggagcga acgacctaca ccgaactgag atacctacag cgtgagctat gagaaagcgc 4380  
 cacgcttccc gaagggagaa aggcggacag gtatccggtg agcggcaggg tcggaacagg 4440  
 agagcgcacg agggagcttc cagggggaaa cgctgggtat ctttatagtc ctgtcgggtt 4500  
 tcgccacctc tgacttgagc gtgatitit gtgatgctcg tcaggggggc ggagcctatg 4560  
 gaaaaacgcc agcaacgcgg cctttttacg gtccctggcc ttgtctggc ctttgctca 4620  
 catgttctt cctgcgttat cccctgattc tgggataac cgtattaccg ccttgagt 4680  
 agctgatacc gctcgcgcga gccgaacgac cgagcgcagc ggtcagtga gcgaggaagc 4740  
 ggaaga 4746

&lt;210&gt; 81

&lt;211&gt; 1213

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence: probe for  
 Southern blot analysis (16SrDNA)

&lt;400&gt; 81

ttatcacggg cagtctgttc agggttccaa actcaacgat ggcaactaaa cacgagggtt 60  
 gcgctcgttg cgggacttaa cccaacacct tacggcacga gctgacgaca gccatgcacc 120  
 acctgtgtcc gcgttccga aggcacccct cctttcaag aggaatcgcg gcatgtcaag 180  
 ccttggttaag gtcttcgct ttgcatcgaa ttaaaccaca tctccaccg ctgtgcggg 240  
 cccccgtcaa ttcttttag ttcatctt gcgaacgtac tcccaggcg ggatactaa 300  
 cgcgttagct acagcactgc acgggtcgat acgcacagcg octagtatcc atcgtttacg 360  
 gctaggacta ctgggtatc taatccatt cgctccccta gtttcgtct ctgagtgtca 420  
 gtgtcggccc agcagagtgc ttgcgcgtt gggttctt cgtatctta cgcattcac 480  
 cgctccaccg gaaattccct ctgccctac cgtactccag ctggtagt tccaccgcct 540  
 gtccagggtt gagccctggg attgacggc ggacttaaaa agccacctac agacgttta 600  
 cgcccaatca ttccggataa cgctgcatc cctgtatta ccgcggctgc tggcacagag 660  
 ttagccgatg cttatcccc agataccgtc atgtctct ctcgggaaa agaagttcac 720  
 gaccgtggg cttctacct ccacgcggca ttgtccgtc aggtttcgc ccattgcgga 780  
 aaattccca ctgtgcctc ccgtaggagt ctggccgtg tctcagccc agtgtggctg 840  
 atcatctct cggaccagt actgatcgc gcttggtta gctattgcct caccaactag 900  
 ctaatcagac gcgagccct cctcggcggt attcctctt ttgtctca gctacgggg 960  
 tattagcagc cgtttccagc tgtgtccc ctccaaggg caggttctta cgcgttactc 1020  
 accgtccgc cactggaaac accactccc gtccgactg catgtgtta gcatgccgc 1080  
 agcgttcatc ctgagccagg atcgaactc ccatgagatt catagtgtca ttactatag 1140  
 ctctctgtt cgtagacaaa gcggattcgg aattgtctt cattocaagg cataactgt 1200  
 atccatgcgc ttc 1213

&lt;210&gt; 82

&lt;211&gt; 949

0817/00028

38

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence: probe for  
Southern blot (23SrDNA)

&lt;400&gt; 82

catcgtcgag agggaaacag cccggatcac cagctaaggc ccctaaatga tcgctcagt 60  
ataaaggagg taggggtgca gagacagcca ggaggttgc ctagaagcag ccaccttga 120  
aagagtgcgt aatagctcac tgatcgagcg ctcttgcgcc gaagatgaac ggggctaagc 180  
gatctgccga agctgtggga tgtaaaaata catcggtagg ggagcgttcc gccttagaga 240  
gaagcctccg cgcgagcggg ggtggacgaa gcggaagcga gaatgtcggc ttgagtaacg 300  
caaacattgg tgagaatcca atgccccgaa aacctagggt ttctctcgca aggttcgtcc 360  
acggagggtg agtcaggggc taagatcagg ccgaaaggcg tagtcgatgg acaacagggtg 420  
aatattcttg tactgcccct tgttggctcc gagggacgga ggaggctagg ttgcccga 480  
gatggttalc ggtcaagaa cgttaagggt ccctgcttgc tcagggttag aagggttaga 540  
gaaatgcct cgagccaatg ttcgaatacc aggcgctacg gcgctgaagt aacctatgcc 600  
atactccag gaaaagctcg aacgacttg agcaagaggg tacctgtacc cgaaccgac 660  
acagggtggg aggtagagaa tacctagggg cgcgagacaa ctctcttaa ggaactcggc 720  
aaaatagccc cgttaactcg ggagaagggg tgctctca caaagggggg cgagtgacc 780  
aggcccgggc gactgttac caaaaacaca ggtctccgca aagtcgtaag accatgtatg 840  
ggggctgacg cctgcccagt gccggaagg caaggaagt ggtgacctga tgacagggga 900  
gccggcgacc gaagccccgg tgaacggcgg ccgtaactat aacggtcct 949

&lt;210&gt; 83

&lt;211&gt; 1287

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence: insert of  
vector pCB456-2

&lt;400&gt; 83

gagctctcag tactcgagac attcacctt agaatagac ttaacttta ctggctaac 60  
ttaaccttg gaccataaaa ggcaccccat tgcgagagt cccttgatta accaaatgaa 120  
acgaagtcta accgaaagca gttatggcaa tggcagaatt ctgacacgg aagatagctt 180  
tggcaaaaaa agcaaaaagc atttaccttg attgagatgt taattgtgtt ggcaattatc 240  
agtattttaa tttgctttt tgtgccaaat ttgatactag agcttcgggt gccaggcggt 300  
gcccttgggc tcccggggcg cgtactcgac gctacctaa gagagtcaag ctaattctaa 360  
tctgcagtct agcgtcggc cgctctagaa ctagtggatc ccccgggctg cagttatttg 420  
ccaactacct tagtgatctc gccttcacg tagtggacaa attcttccaa ctgatctgcg 480  
cgcgaggcca agcgatctc ttctgtcca agataagcct gtctagcttc aagtatgacg 540  
ggctgatact gggccggcag gcgctccat gccagtcgg cagcgacatc cttcggcgcg 600  
atttggccg ttactgcgt gtaccaaag cgggacaacg taagcactac atttcgtca 660  
tcgccagccc agtcggggcg cgagttccat agcgtaagg ttcatitag cgcctcaaat 720  
agatcctgt cagggaaccgg atcaaagagt tctccgccg ctggacctac caaggcaacg 780  
ctatgtctc ttgctttgt cagcaagata gccagatcaa tgcgatcgt ggctggctcg 840  
aagatacctg caagaatgtc attgcgtgc cattctcaa attgcagtic gcgcttagct 900

0817/00028

39

ggataacgcc acggaatgat gtcgtcgtgc acaacaatgg tgacttctac agcgcggaga 960  
 atctcgtctt ctccagggga agccgaagtt tccaaaaggt cgttgatcaa agctcgccgc 1020  
 gttgttcat caagccttac ggtcaccgta accagcaaat caatatcact gtgtggcttc 1080  
 agggccgcat cactgctgga gccgtacaaa tgtacggcca gcaacgtcgg ttgagatgg 1140  
 cgtcgtatga cgccaactac ctctgatagt tgagttgata cttcggcgat aaccgcttca 1200  
 cgagccatga gatcctccag atccatgtat cattatagat aattgaagag tgaatgtcaa 1260  
 gtcgacctcg agggggggcc cgggtacc 1287

&lt;210&gt; 84

&lt;211&gt; 1387

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> Description of the artificial sequence: insert of  
 vector pCB528-2

&lt;400&gt; 84

ggtaacccag ttaactggag atctcgagga ggtcacatgg gagcttggat tgaacaagat 60  
 ggattgcacg caggtttcc ggccgcttgg gtggagaggc taltcggcta tgactgggca 120  
 caacagacaa tcggctgctc tgatgccgcc gtgtccggc tgcagcga ggggcggccg 180  
 gttcttttg tcaagaccga cctgtcgggt gccctgaalg aactgcagga cgaggcagcg 240  
 cggctatcgt ggctggccac gacgggcgtt ccttgcgcag ctgtgctcga cgttgcact 300  
 gaagcgggaa gggactggct gctattgggc gaagtgccgg ggcaggatct cctgtcatct 360  
 caccttgctc ctgccgagaa agtatccatc atggctgatg caatgcggcg gctgcatacg 420  
 cttgatccgg ctacctgccc attcgaccac caagcgaaac atgcacga gcgagcacgt 480  
 actcggatgg aagccggctt tgcgatcag gatgatctgg acgaagagca tcaggggctc 540  
 gcgccagccg aactgttcgc caggctcaag gcgcgcacgc ccgacggcga ggatctcgtc 600  
 gtgacccatg gcgatgctg ctgccgaat atcatgggtg aaaatggccg cttttctgga 660  
 ttcatcgact gtggccggct ggggtgtggc gaccgctatc aggacatagc gttggctacc 720  
 cgtgatattg ctgaagagct tggcggcgaa tgggctgacc gcttcctcgt gcittacggt 780  
 atcggcgtc ccgattcgca gcgcacgcc ttctatgcc ttcttgacga gticttctga 840  
 tctagatagg aggtcatcat ggcgctcacc aatgcicaaa tcttggctgt gattgacagt 900  
 tgggaagaaa cagtgggtca gtttcagtg ataacgcacc atgtaccatt aggtggcgg 960  
 ctgcaaggaa cgctccattg ttacgagatc cccctagcag ctccctatgg ggttggcttt 1020  
 gctaagaatg ggccatcccg ctggcaatac aaacggacaa tcaatcaagt cgtccacaga 1080  
 tgggggtccc acacagtccc ttictatta gaaccggata acatcaacgg caaacctgc 1140  
 acagcatcgc acctatgtca taatactga tgccacaatc ccttgcactt gtgctgggag 1200  
 tcaatagacg acaacaaagg cagaaactgg tgcccgggtc ccaacggggg atgtgtccat 1260  
 gcggtggttt gtttaaggca ggtccgttg tacggccag gggcgactgt ggcaggtcct 1320  
 caacaaaggg gcagtcactt tgtgtataa ctgcagaagc ttaagccag ttaactgggc 1380  
 ggagctc 1387

We claim:

1. A method for the integration of a DNA sequence into the  
5 plastid DNA of a multi-celled plant or cell derived therefrom  
and for the selection of predominantly homotransplastomic  
cells or plants, wherein
  - a) the plastid DNA molecules of said multi-celled plant  
10 comprise at least one recognition sequence for the  
directed induction of DNA double-strand breaks and
  - b) at least one enzyme suitable for the induction of DNA  
15 double-strand breaks at the recognition sequence for the  
directed induction of DNA double-strand breaks and at  
least one transformation construct comprising an  
insertion sequence are combined in at least one plastid  
of said multi-celled plant or cell derived therefrom, and
  - 20 c) DNA double-strand breaks are induced at the recognition  
sequences for the directed induction of DNA double-strand  
breaks, and
  - 25 d) the insertion sequence inserts into the plastid DNA, the  
functionality of the recognition sequence for the  
directed induction of DNA double-strand breaks being  
deactivated so that said recognition sequence is no  
longer capable of being cleaved by the enzyme suitable  
30 for the induction of DNA double-strand breaks, and
  - e) plants or cells in which the insertion sequence has been  
inserted into the plastid DNA molecules are isolated.
- 35 2. A method according to claim 1, wherein the recognition  
sequence for the directed induction of DNA double-strand  
breaks and the insertion sequence are flanked at least  
unilaterally by sequences with sufficient length and  
sufficient homology to ensure homologous recombination with  
40 each other.
3. A method according to either of claims 1 and 2, wherein the  
transformation construct encompasses at least one of the  
elements selected from the group consisting of

- 5  
10  
15  
20
- i) expression cassette for an enzyme suitable for the induction of DNA double-strand breaks at the recognition sequence for the directed induction of DNA double-strand breaks
  - ii) positive selection markers
  - iii) negative selection markers
  - iv) reporter genes
  - v) replication origins
  - vi) multiple cloning regions
  - vii) border sequences for agrobacterium transfection
  - viii) sequences which make possible a homologous recombination or insertion into the genome of a host organism.
- 25  
30
4. A method according to any of claims 1 to 3, wherein an enzyme suitable for the induction of DNA double-strand breaks at the recognition sequence for the directed induction of DNA double-strand breaks is selected from the group consisting of restriction endonucleases, homing endonucleases, chimeric nucleases, mutated restriction or homing endonucleases and RNA protein particles derived from mobile group II introns, and fusion proteins of the abovementioned proteins with plastid localization sequences.
- 35  
40  
45
5. A method according to any of claims 1 to 4, wherein the enzyme suitable for the induction of DNA double-strand breaks at the recognition sequence for the directed induction of DNA double-strand breaks is selected from the group of the homing endonucleases consisting of F-SceI, F-SceII, F-SuvI, F-TevI, F-TevII, I-AmaI, I-AniI, I-CeuI, I-CeuAIIP, I-ChuI, I-Cmoel, I-CpaI, I-CpaII, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIVP, I-CsmI, I-CvuI, I-CvuAIP, I-DdiII, I-DirI, I-DmoI, I-HspNIP, I-LlaI, I-MsoI, I-NaaI, I-NanI, I-NclIP, I-NgrIP, I-NitI, I-NjaI, I-Nsp236IP, I-PakI, I-PboIP, I-PcuIP, I-PcuAI, I-PcuVI, I-PgrIP, I-PobIP, I-PorI, I-PorIIP, I-PpbIP, I-PpoI, I-SPBetaIP, I-ScaI, I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-SexIP, I-SneIP, I-SpomCP, I-SpomIP, I-SpomIIP, I-SquIP, I-Ssp6803I, I-SthPhiJP, I-SthPhiST3P, I-SthPhiS3bP, I-TdeIP, I-TevI, I-TevII, I-TevIII, I-UarAP, I-UarHGPAIP,

- I-UarHGPA13P, I-VinIP, I-ZbiIP, PI-MtuI, PI-MtuHIP,  
PI-MtuHIIP, PI-PfuI, PI-PfuII, PI-PkoI, PI-PkoII, PI-PspI,  
PI-Rma43812IP, PI-SPBetaIP, PI-SceI, PI-TfuI, PI-TfuII,  
PI-ThyI, PI-TliI, PI-TliII and fusion proteins of the above  
5 with plastid localization sequences.
6. A method according to any of claims 1 to 5, wherein the  
enzyme suitable for the induction of DNA double-strand breaks  
at the recognition sequence for the directed induction of DNA  
10 double-strand breaks is selected from the group of the homing  
endonucleases consisting of the enzymes as shown in SEQ ID  
NO: 5, 12, 14 and 71.
7. A method according to any of claims 1 to 5, wherein the  
15 enzyme suitable for the induction of DNA double-strand breaks  
at the recognition sequence for the directed induction of DNA  
double-strand breaks is expressed using an expression  
cassette encompassing a nucleic acid sequence encoding said  
enzyme.  
20
8. An expression cassette comprising an enzyme suitable for the  
induction of DNA double-strand breaks at the recognition  
sequence for the directed induction of DNA double-strand  
breaks as defined in any of claims 4 to 6 under the control  
25 of a promoter which is functional in plant plastids.
9. An expression cassette comprising a fusion protein of a  
plastid translocation sequence and an enzyme suitable for the  
induction of DNA double-strand breaks at the recognition  
30 sequence for the directed induction of DNA double-strand  
breaks as defined in any of claims 4 to 6 under the control  
of a promoter which is functional in the plant nucleus.
10. A multi-celled plant obtained by a method according to any of  
35 claims 1 to 7.
11. A multi-celled plant comprising an expression cassette  
according to claim 8 inserted into the plastid DNA.
- 40 12. A multi-celled plant comprising an expression cassette  
according to claim 9 inserted into the nuclear DNA.
13. A multi-celled plant according to either of claims 10 and 12  
45 selected from the group consisting of Arabidopsis thaliana,  
tobacco, Tagetes, wheat, rye, barley, oats, oilseed rape,

## 116

maize, potato, sugar beet, soybean, sunflower, pumpkin/squash or peanut.

14. A cell culture, organ, tissue, part or transgenic propagation material derived from a multi-celled plant according to claims 10 to 13.
15. The use of an organism according to any of claims 10 to 13 or of cell cultures, organs, tissues, parts or transgenic propagation material derived therefrom according to claim 14 as food, feed or seed or for the preparation of pharmaceuticals or fine chemicals.
16. A DNA construct comprising at least one nucleic acid sequence and intron sequence element which are capable of ensuring, in a ribonucleic acid sequence derived from said DNA construct, the deletion of the ribonucleic acid fragment encoding said nucleic acid sequence, said nucleic acid sequence being heterologous with regard to said intron sequence elements.
17. A DNA construct according to claim 16, wherein the nucleic acid sequence is flanked by at least one splice acceptor sequence and a splice donor sequence.
18. A DNA construct according to claim 16 or 17, wherein the DNA construct comprises, at the 5' and 3' end, sequences H1 and H2, respectively, with a sufficient length and homology with plastid sequences H1' and H2', respectively, to ensure homologous recombination between H1 and H1', or H2 and H2', and thus insertion of the H1- and H2-flanked sequence into the plastome.
19. A transgenic plasmid DNA comprising at least one nucleic acid sequence and intron sequence elements which are capable of ensuring, in a ribonucleic acid sequence derived from said transgenic plasmid DNA construct, the deletion of the ribonucleic acid fragment encoding said nucleic acid sequence, said nucleic acid sequence being heterologous with regard to said intron sequence elements.
20. A transgenic plasmid DNA according to claim 19, wherein the nucleic acid sequence is flanked by at least one splice acceptor sequence and a splice donor sequence.
21. A transgenic plant comprising a transgenic plasmid DNA according to claim 20.



Application number / numéro de demande: EP02/14302

Figures: 9/11/12

Pages: \_\_\_\_\_

Unscannable items  
received with this application  
(Request original documents in File Prep. Section on the 10<sup>th</sup> floor)

Documents reçu avec cette demande ne pouvant être balayés  
(Commander les documents originaux dans la section de préparation des dossiers au  
10<sup>ème</sup> étage)

---

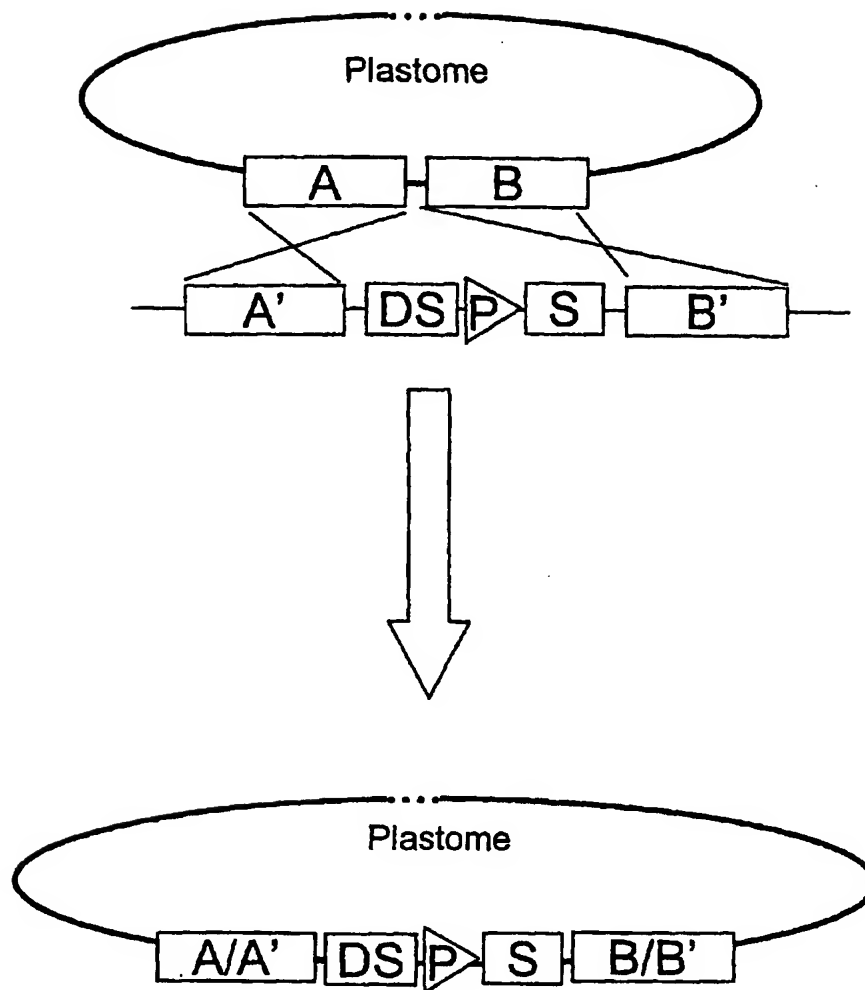


Fig. 1

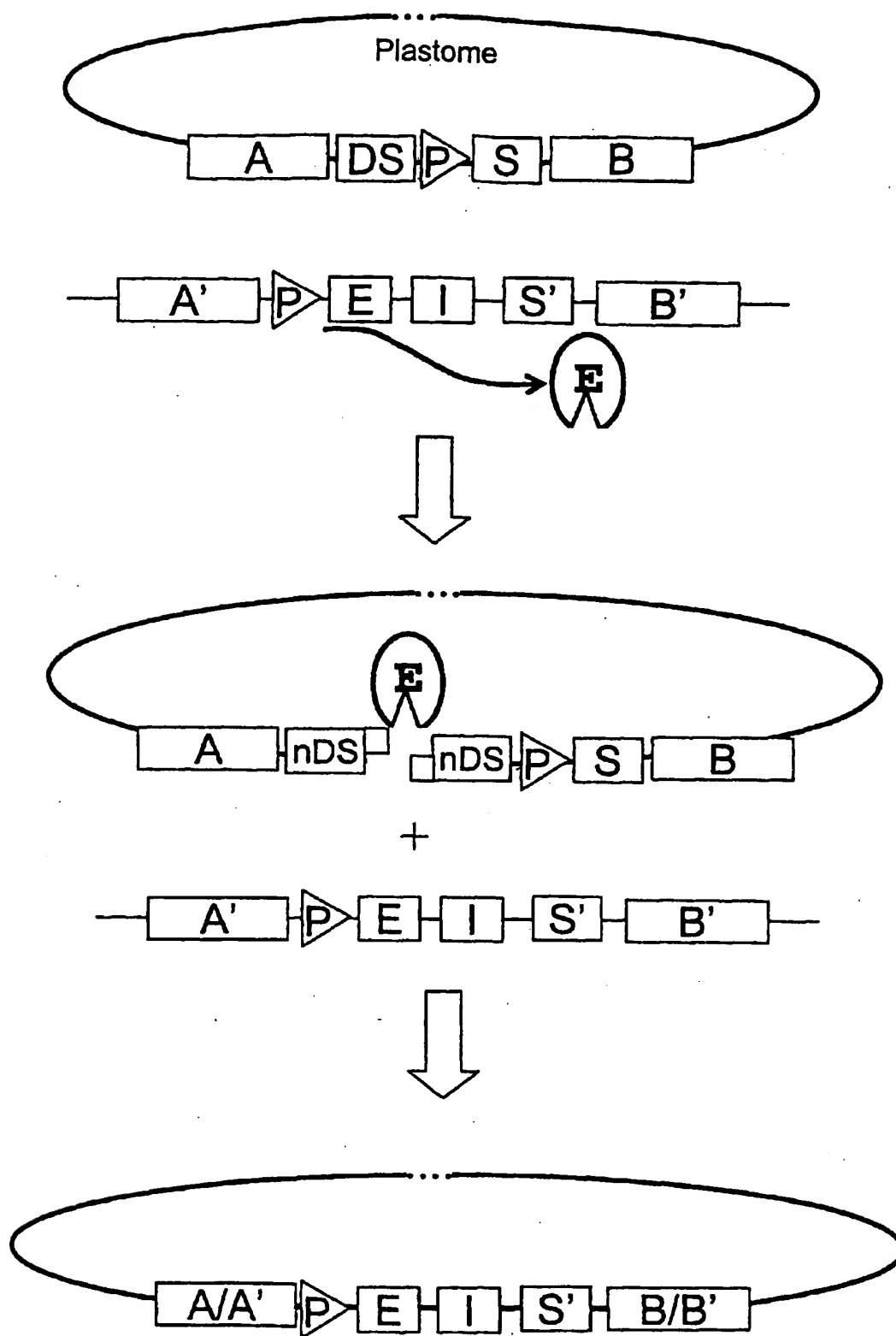


Fig. 2A

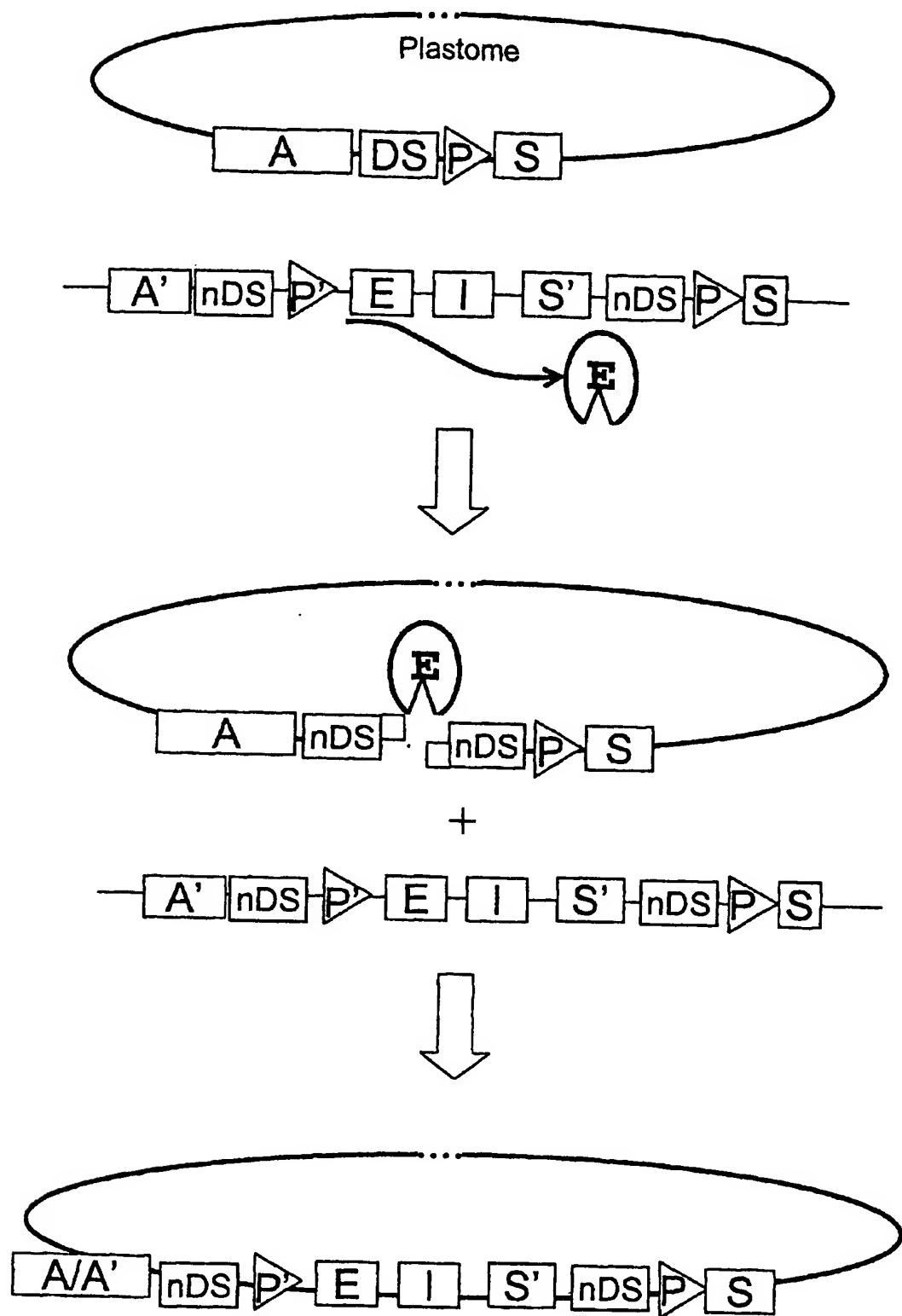


Fig. 2B

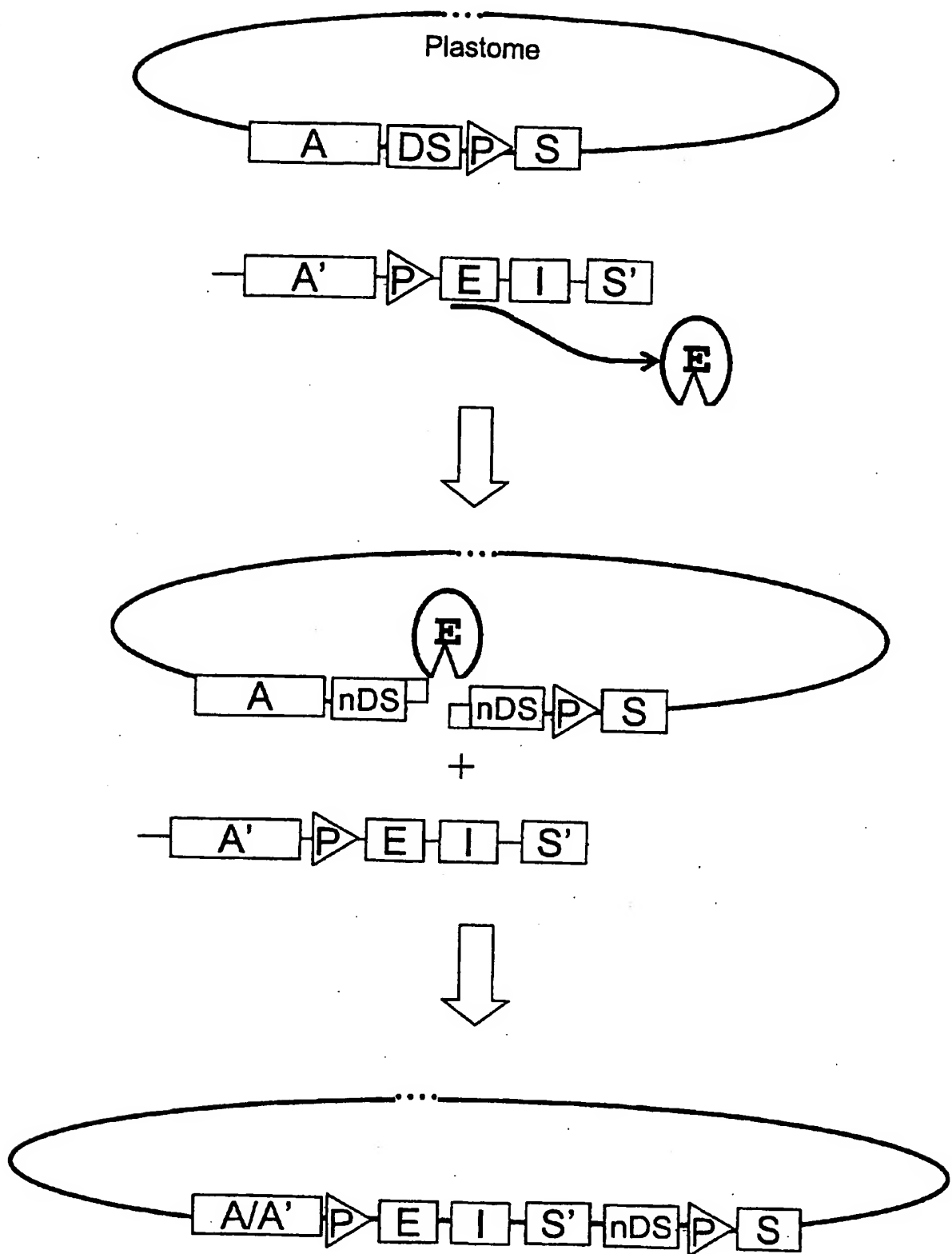


Fig. 2C

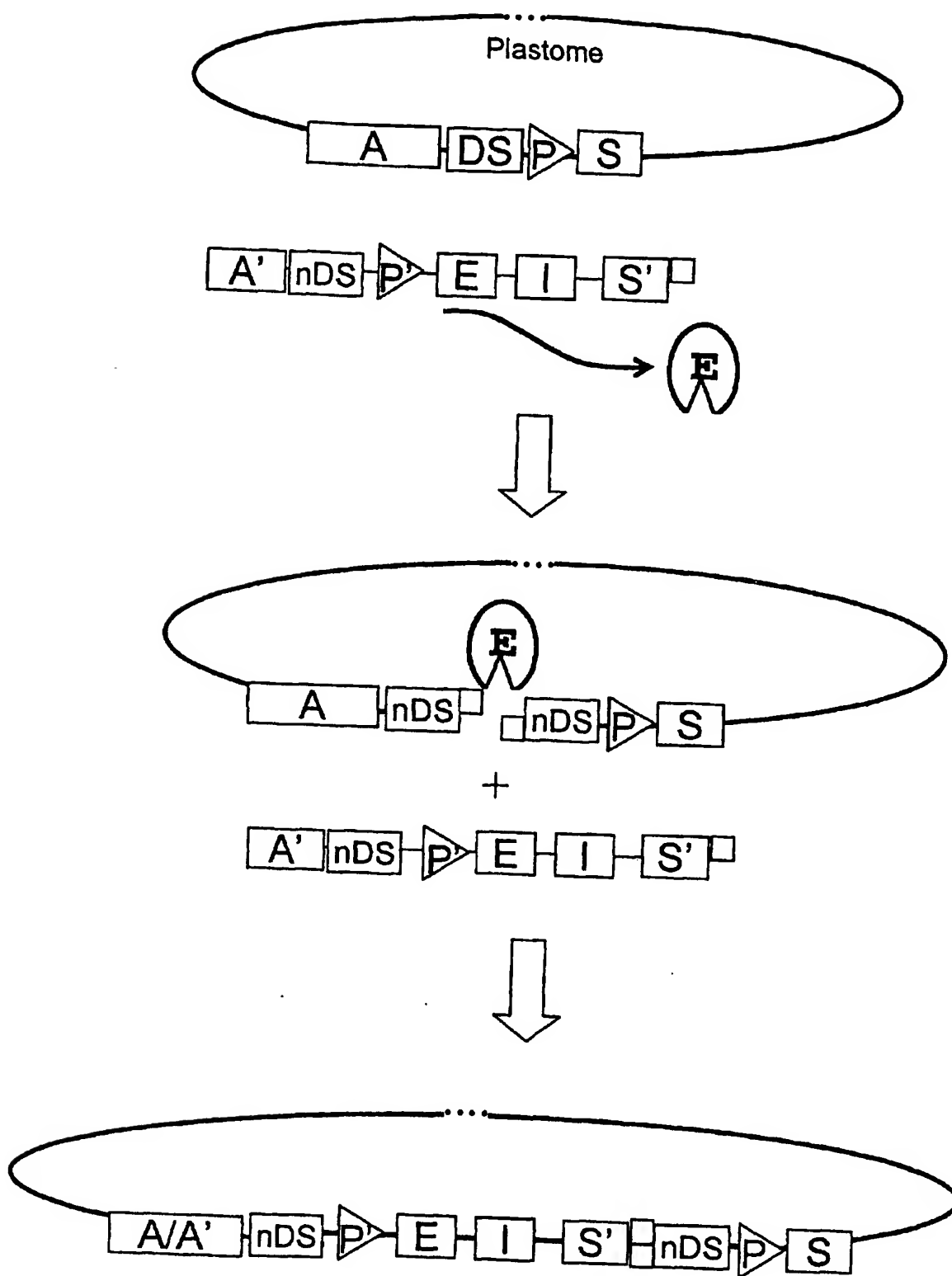


Fig. 2D

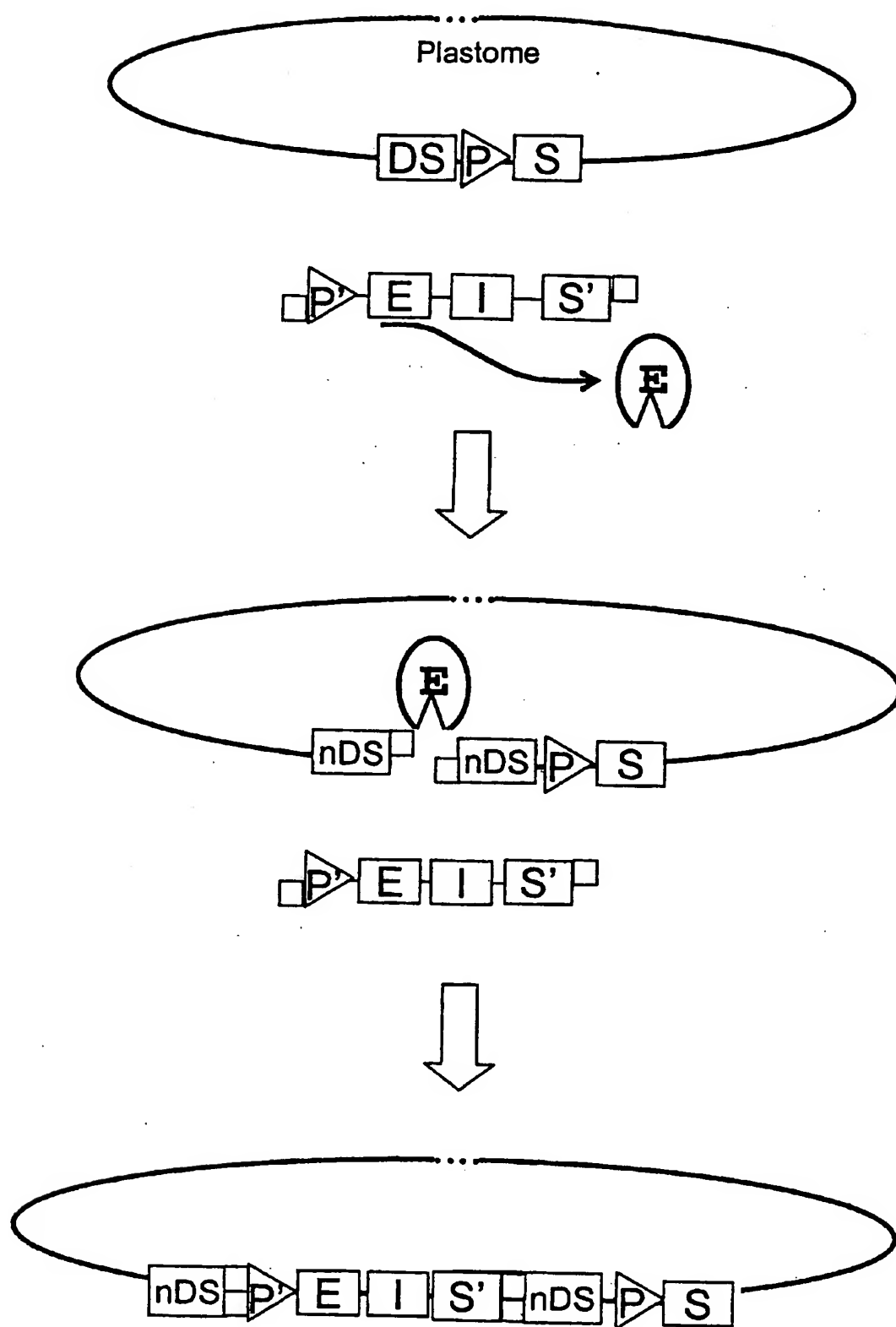


Fig. 2E

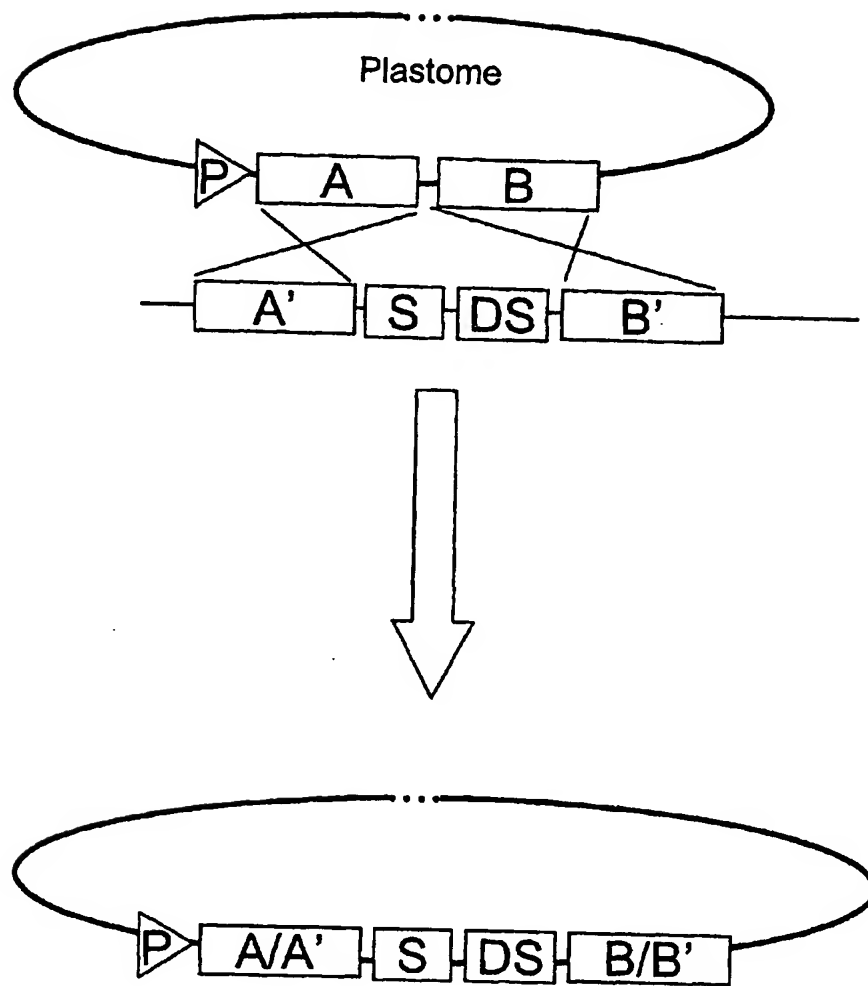


Fig. 3



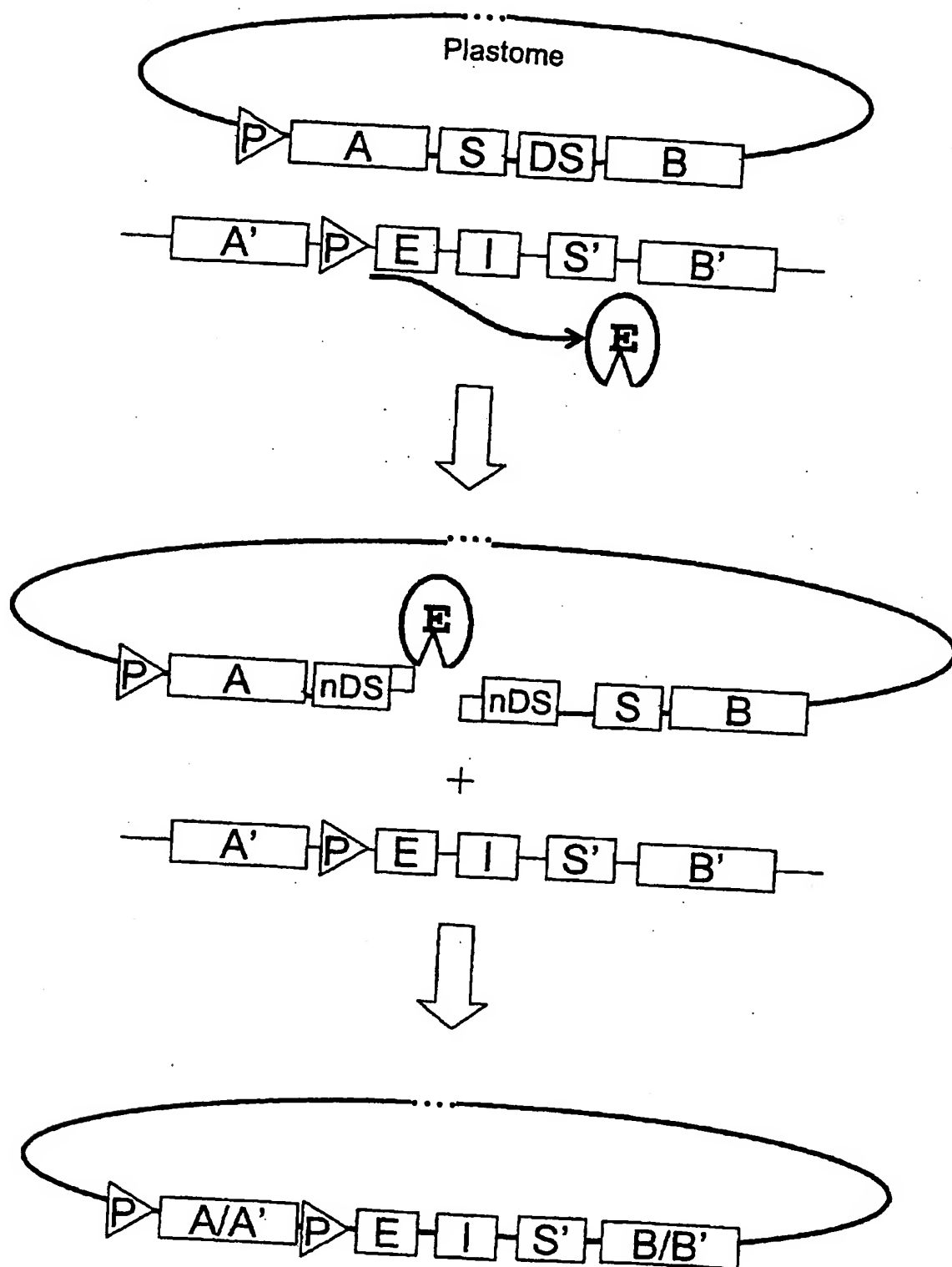


Fig. 4A

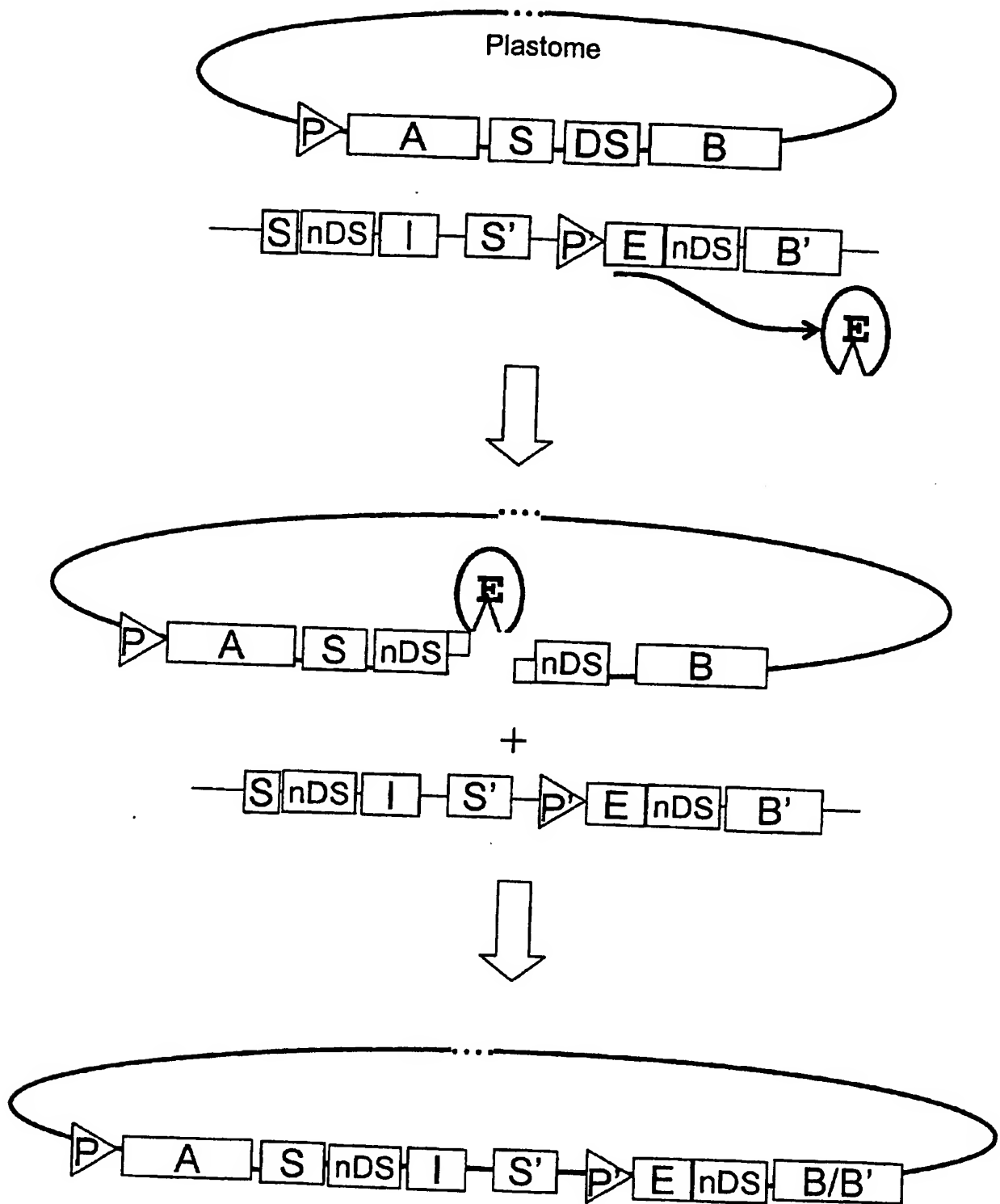


Fig. 4B

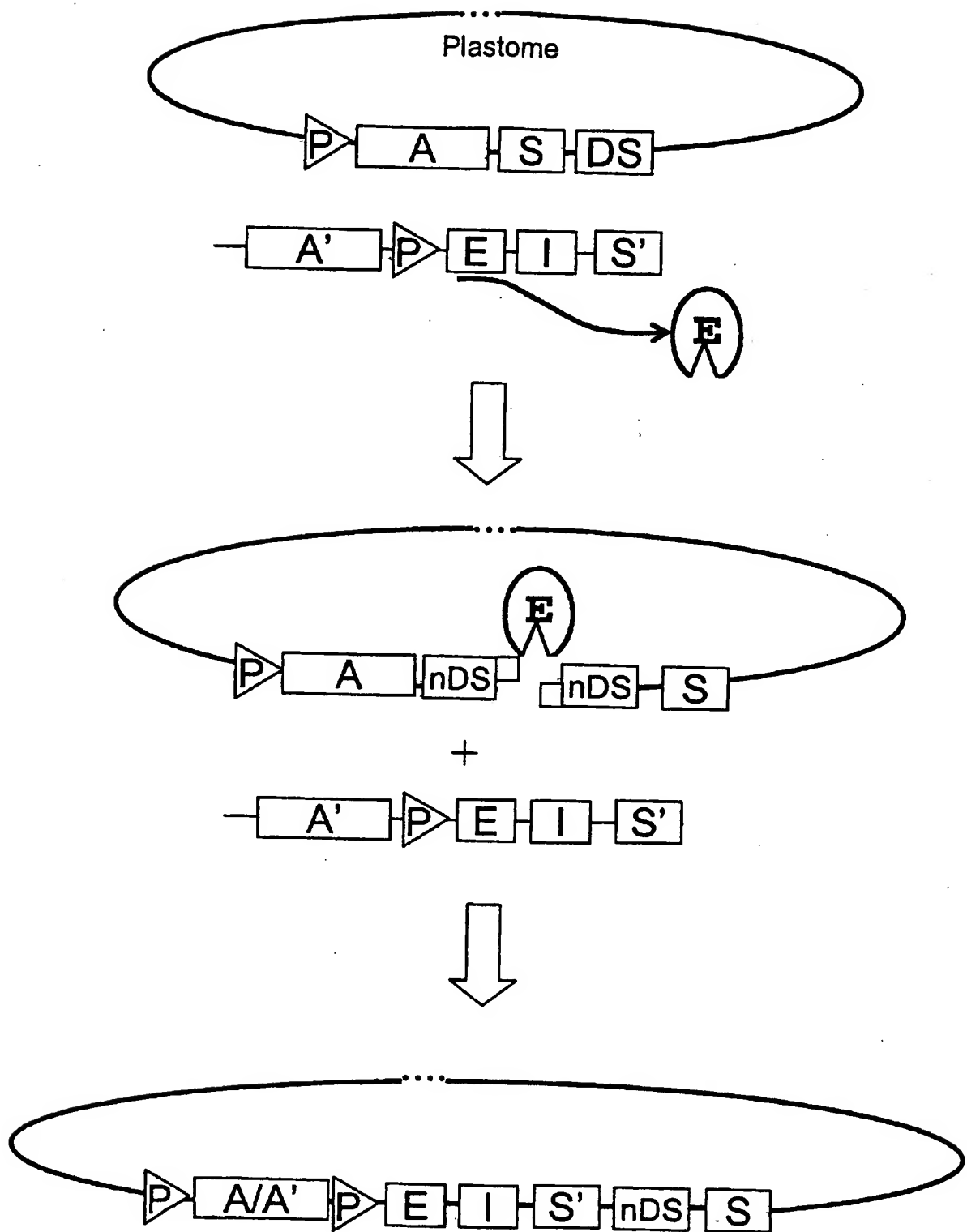


Fig. 4C

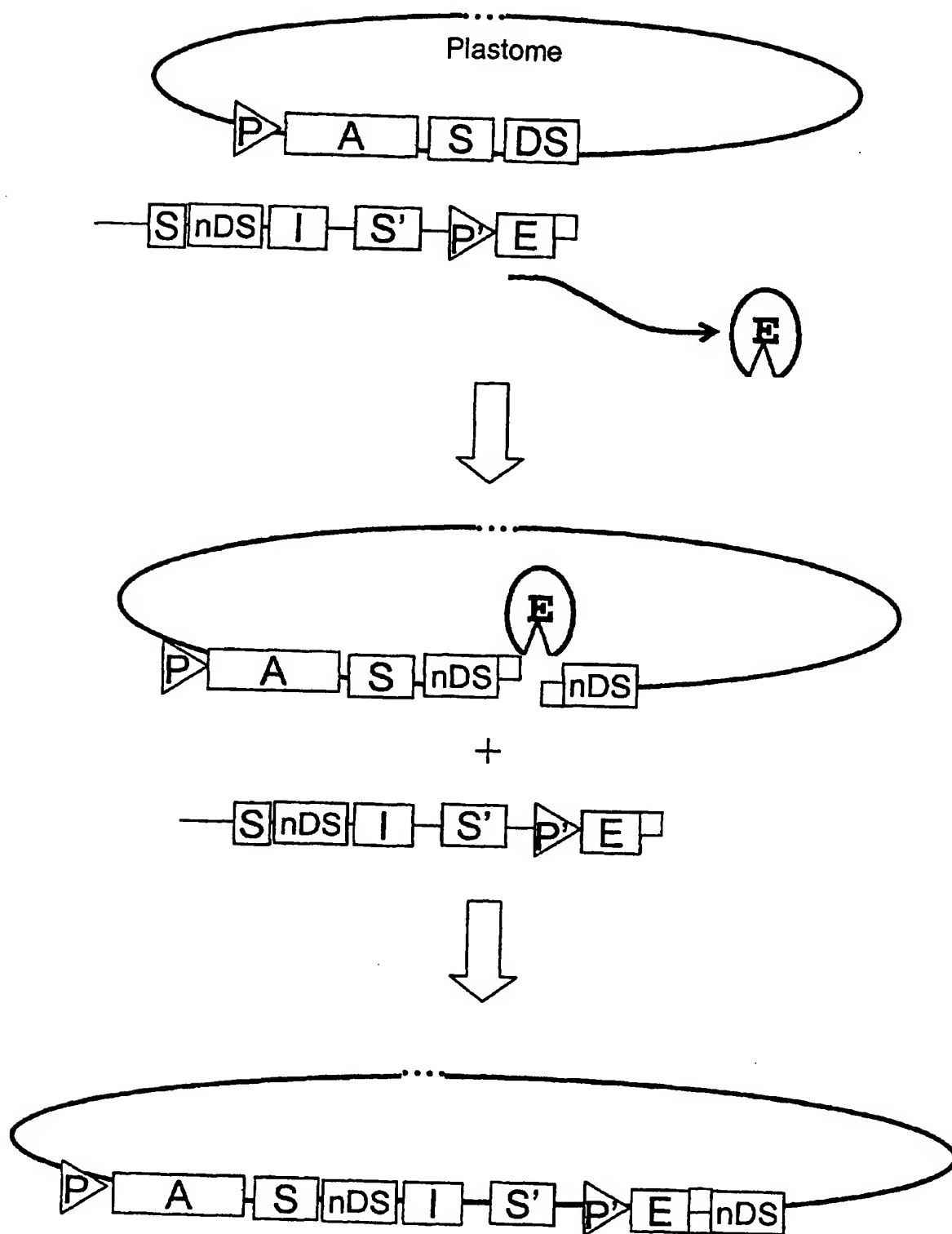


Fig. 4D

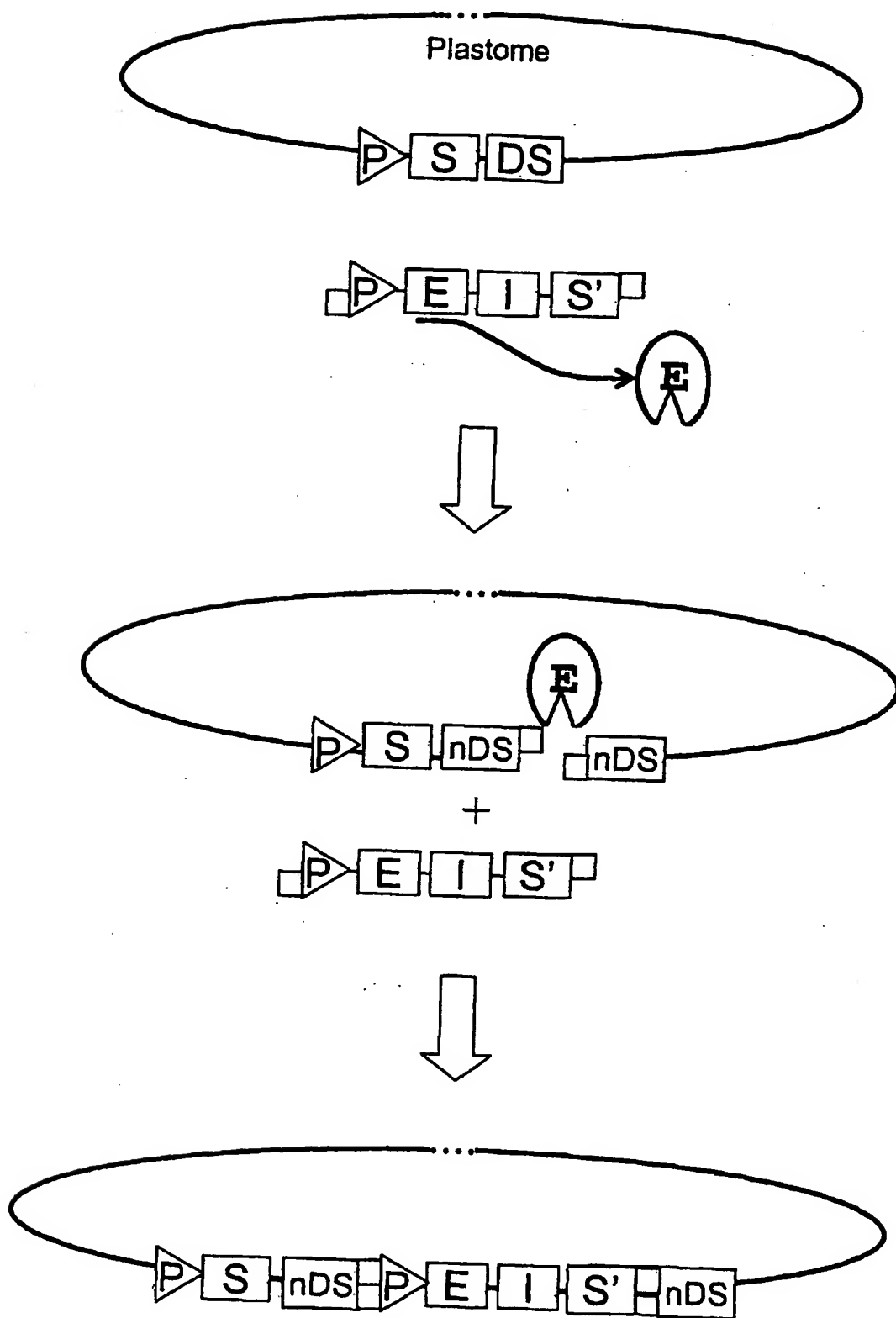


Fig. 4E

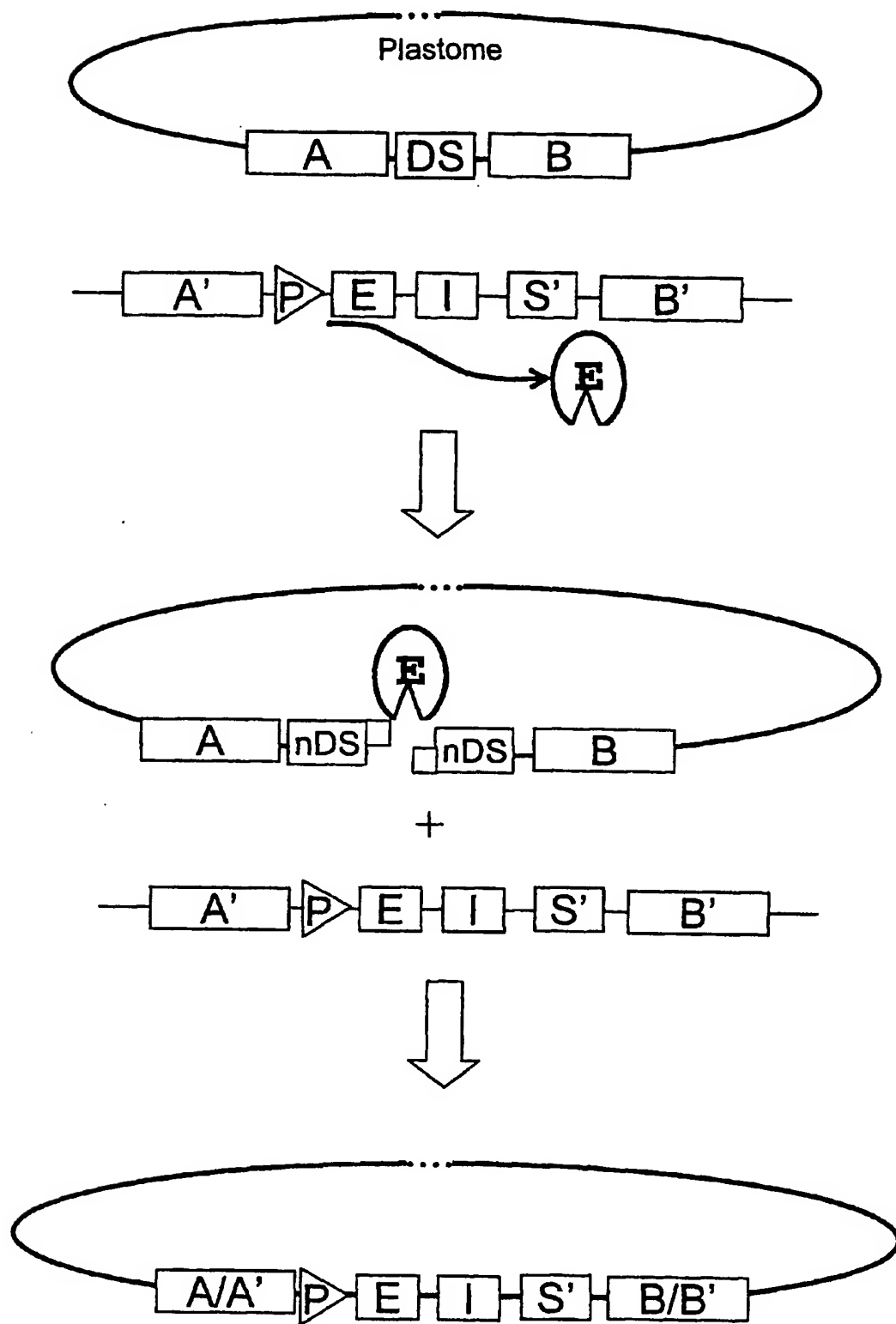


Fig. 5A

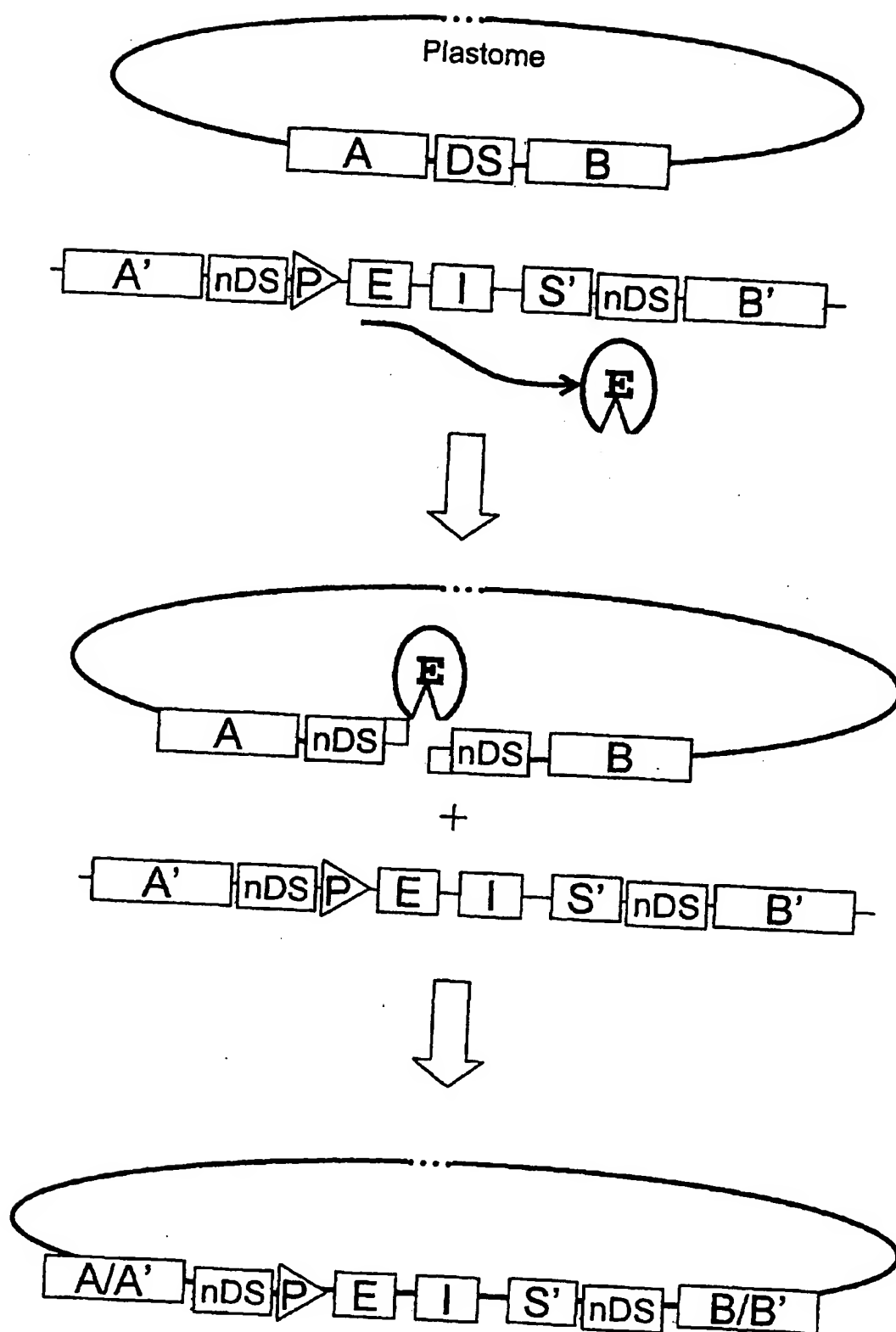


Fig. 5B

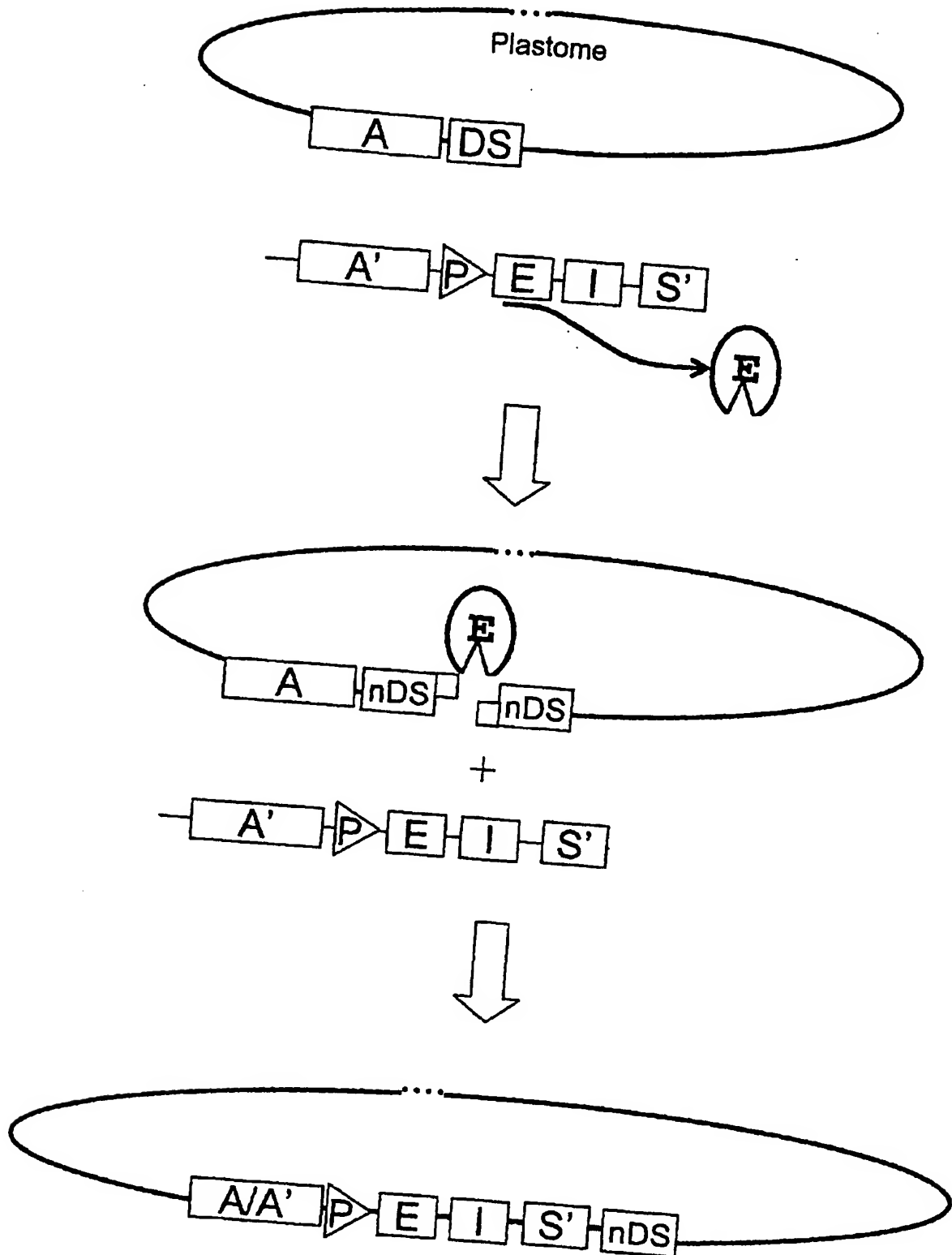


Fig. 5C



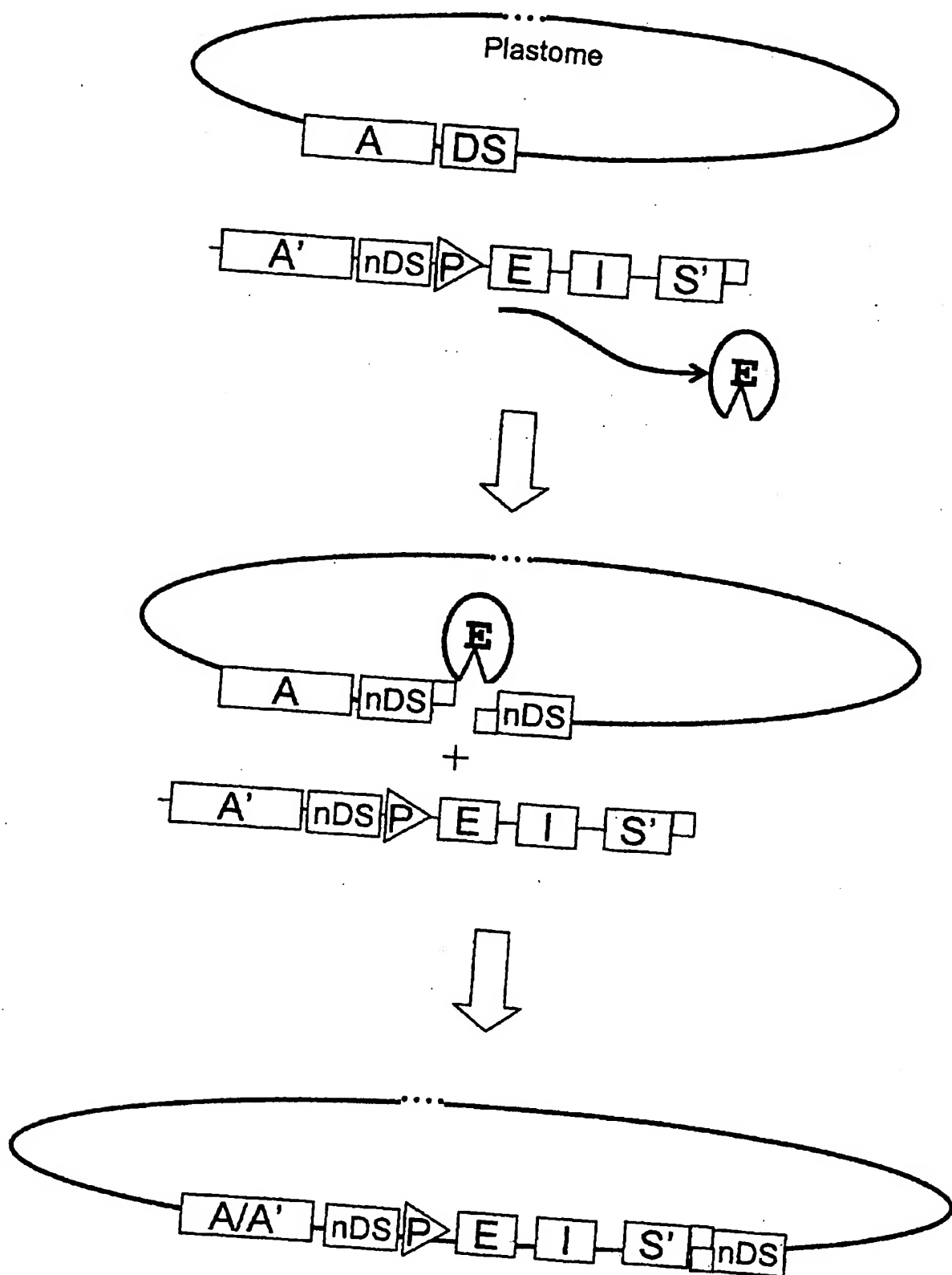


Fig. 5D

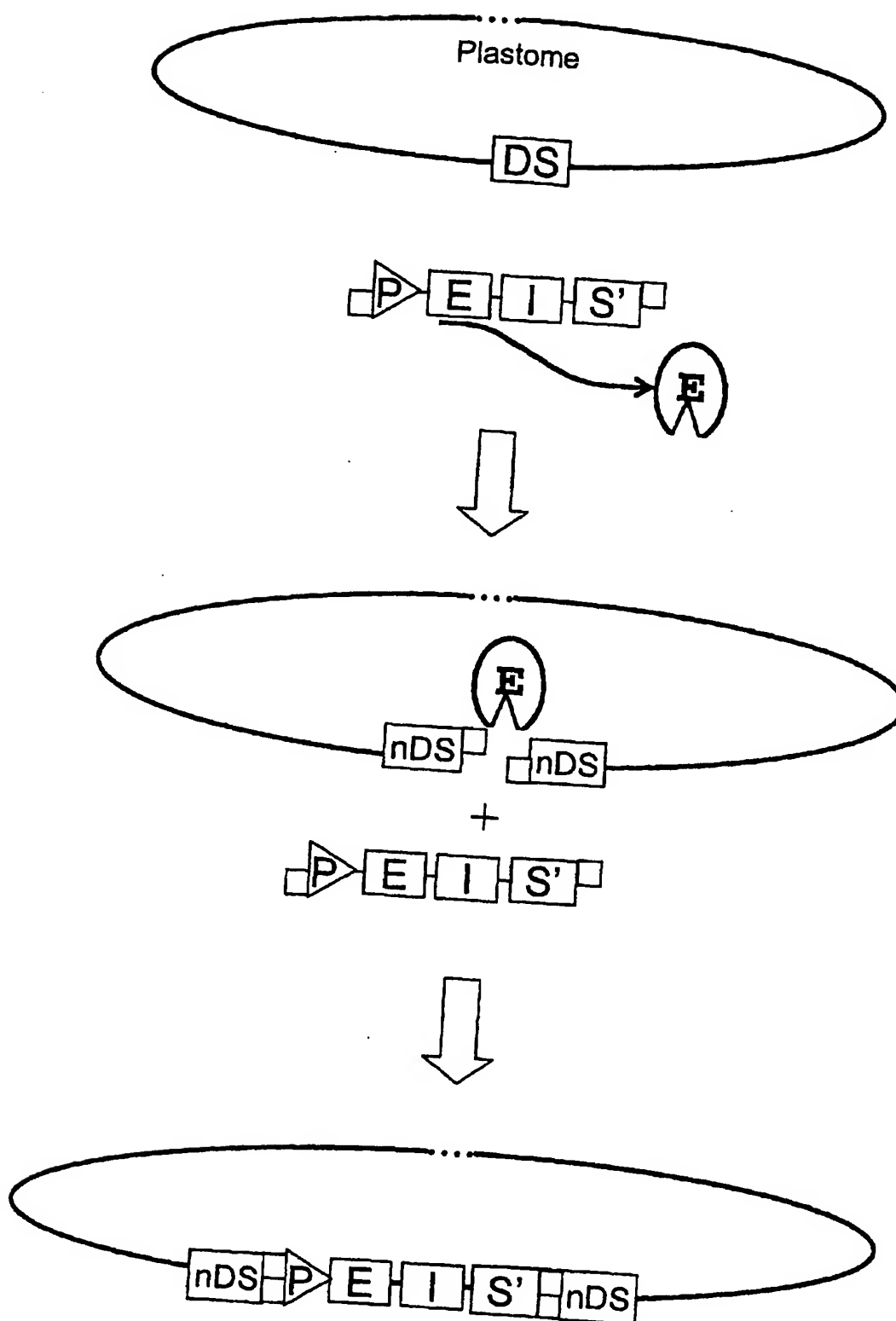


Fig. 5E

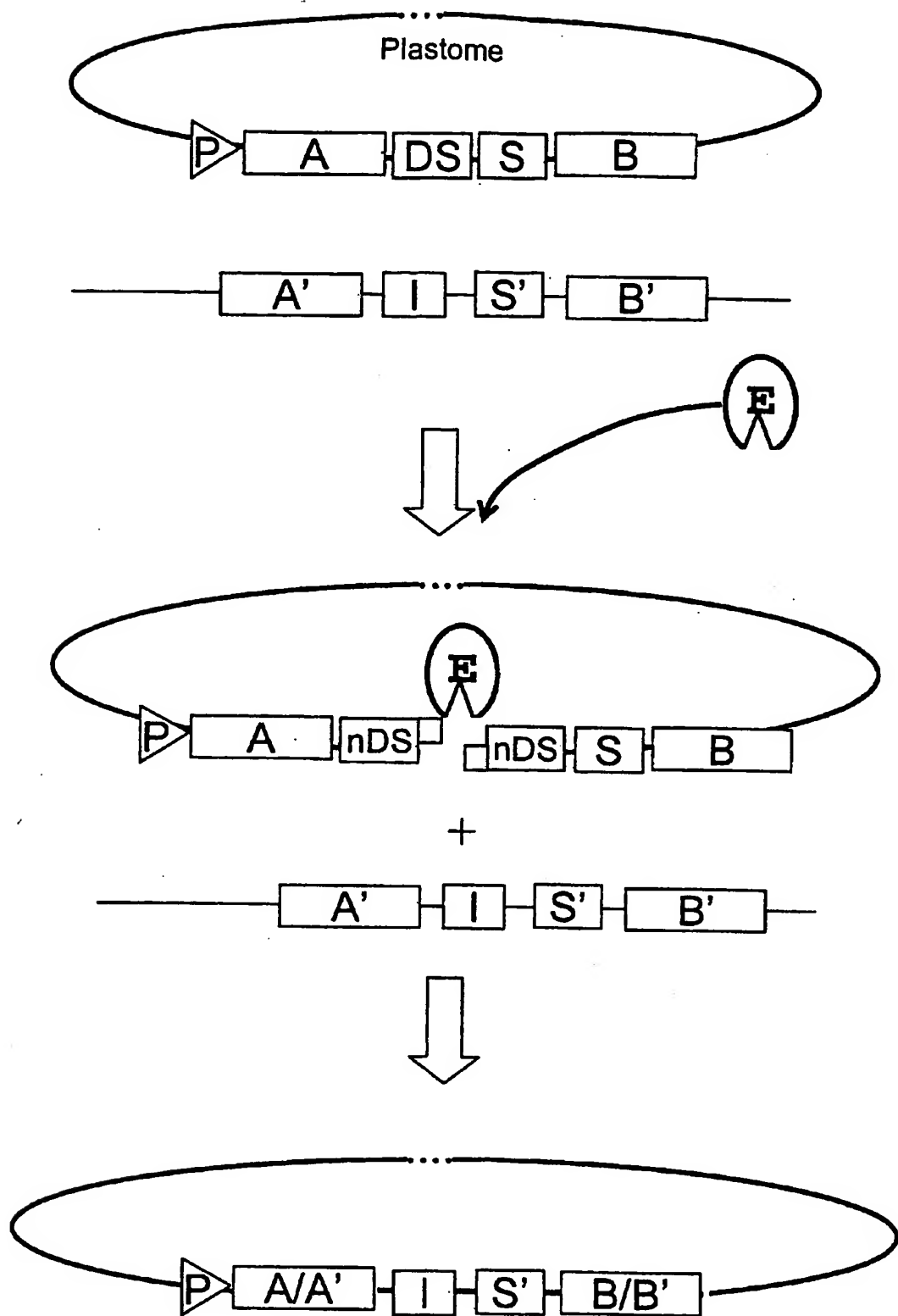


Fig. 6A

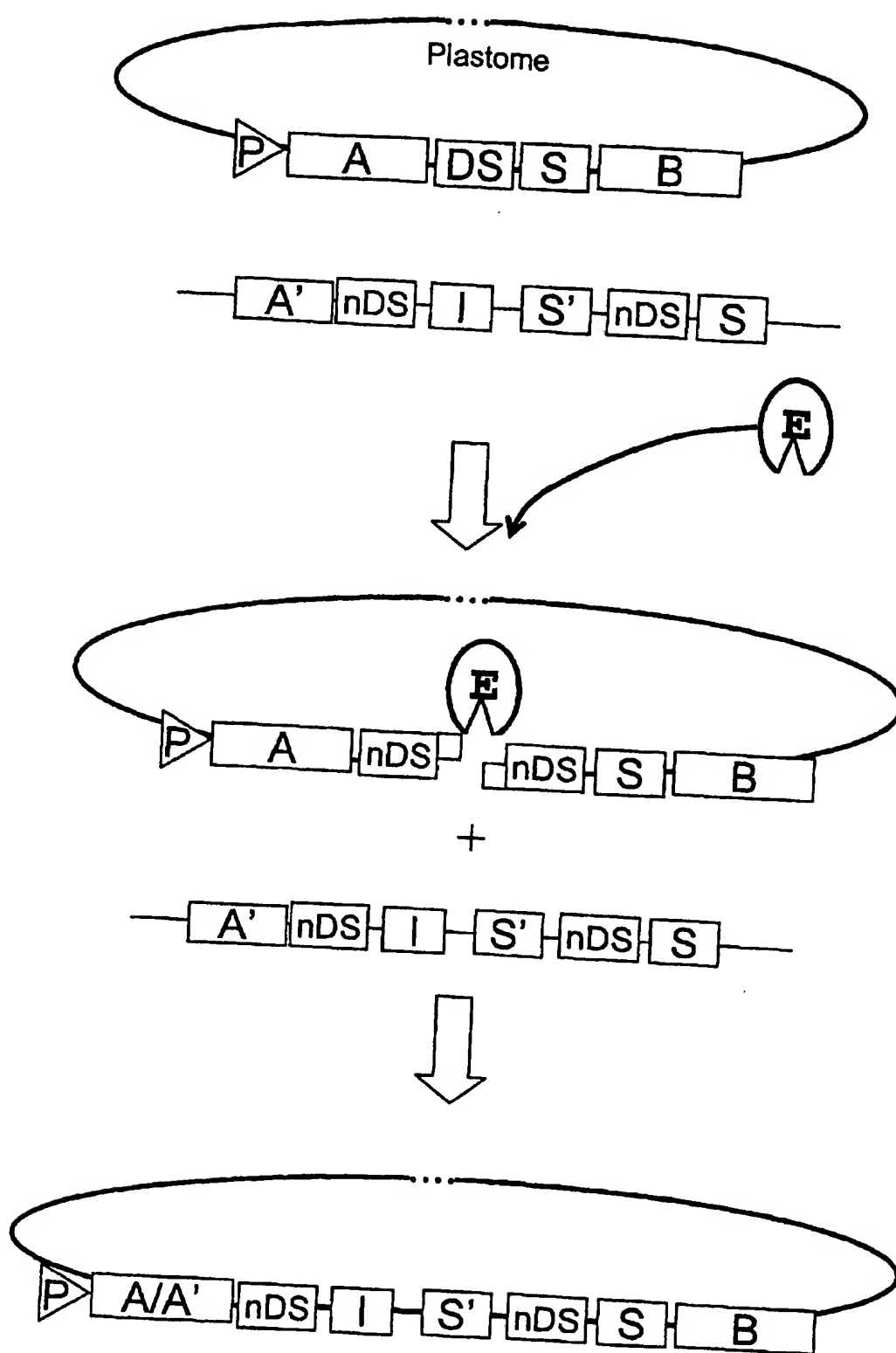


Fig. 6B

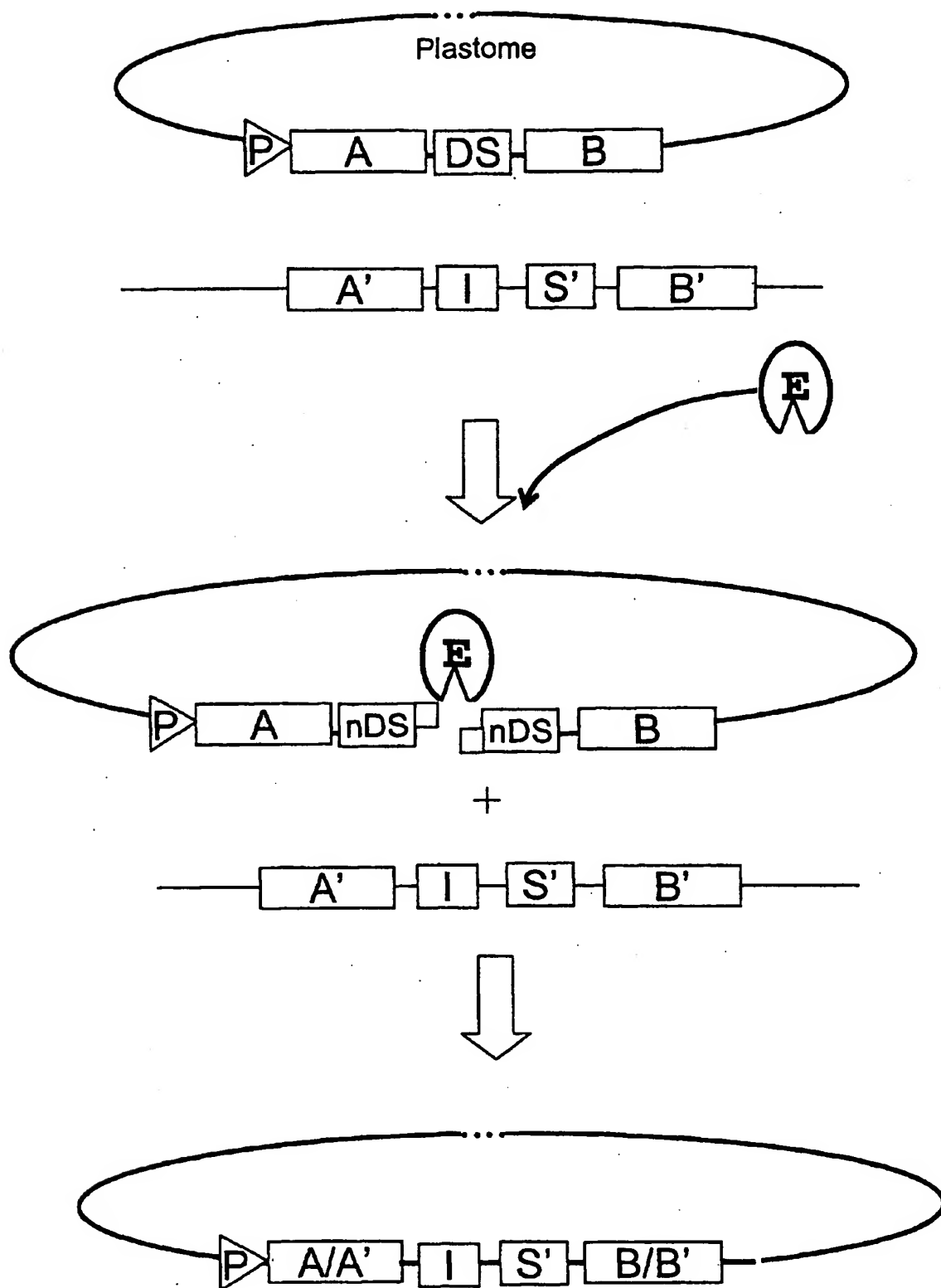


Fig. 6C

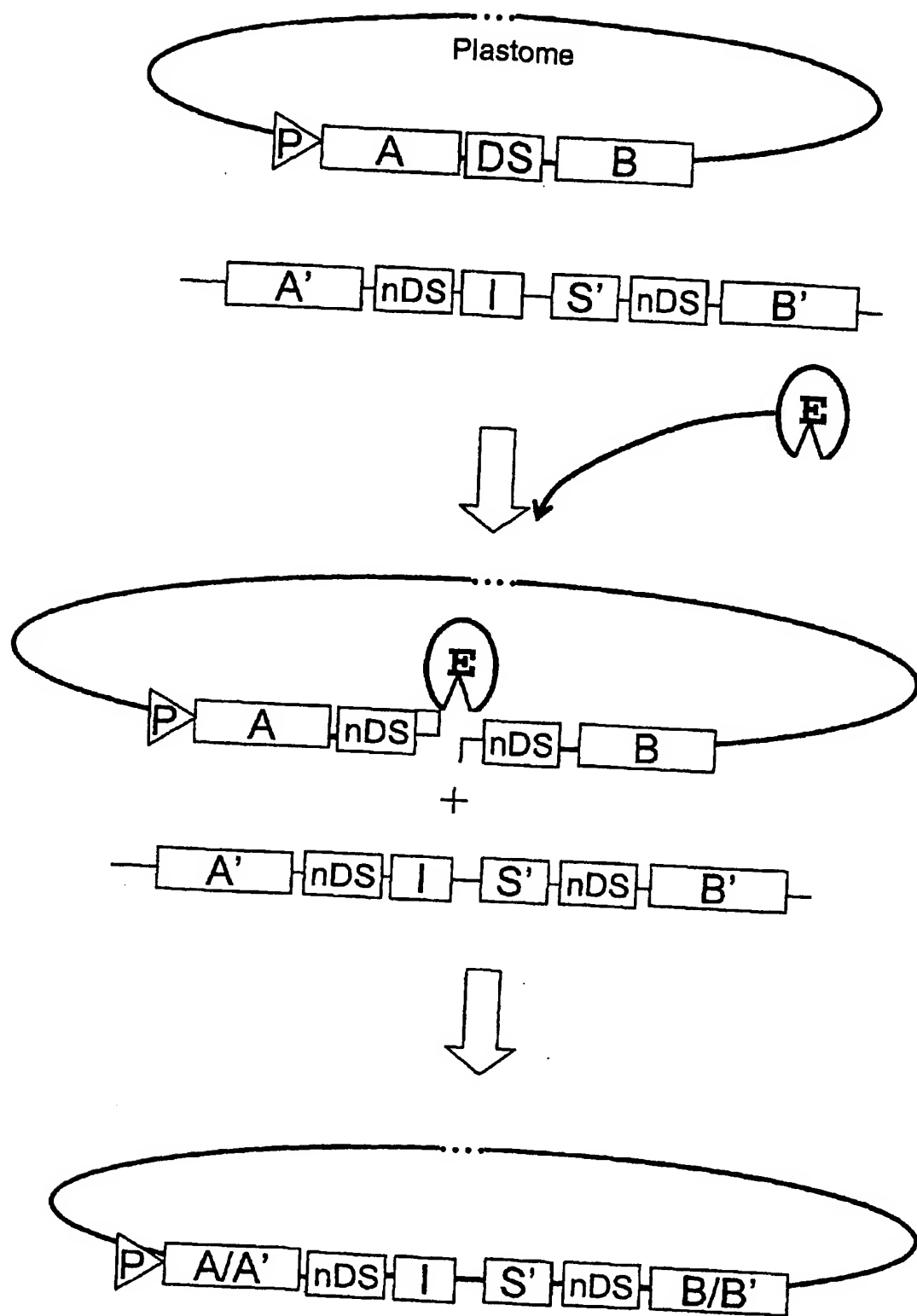


Fig. 6D

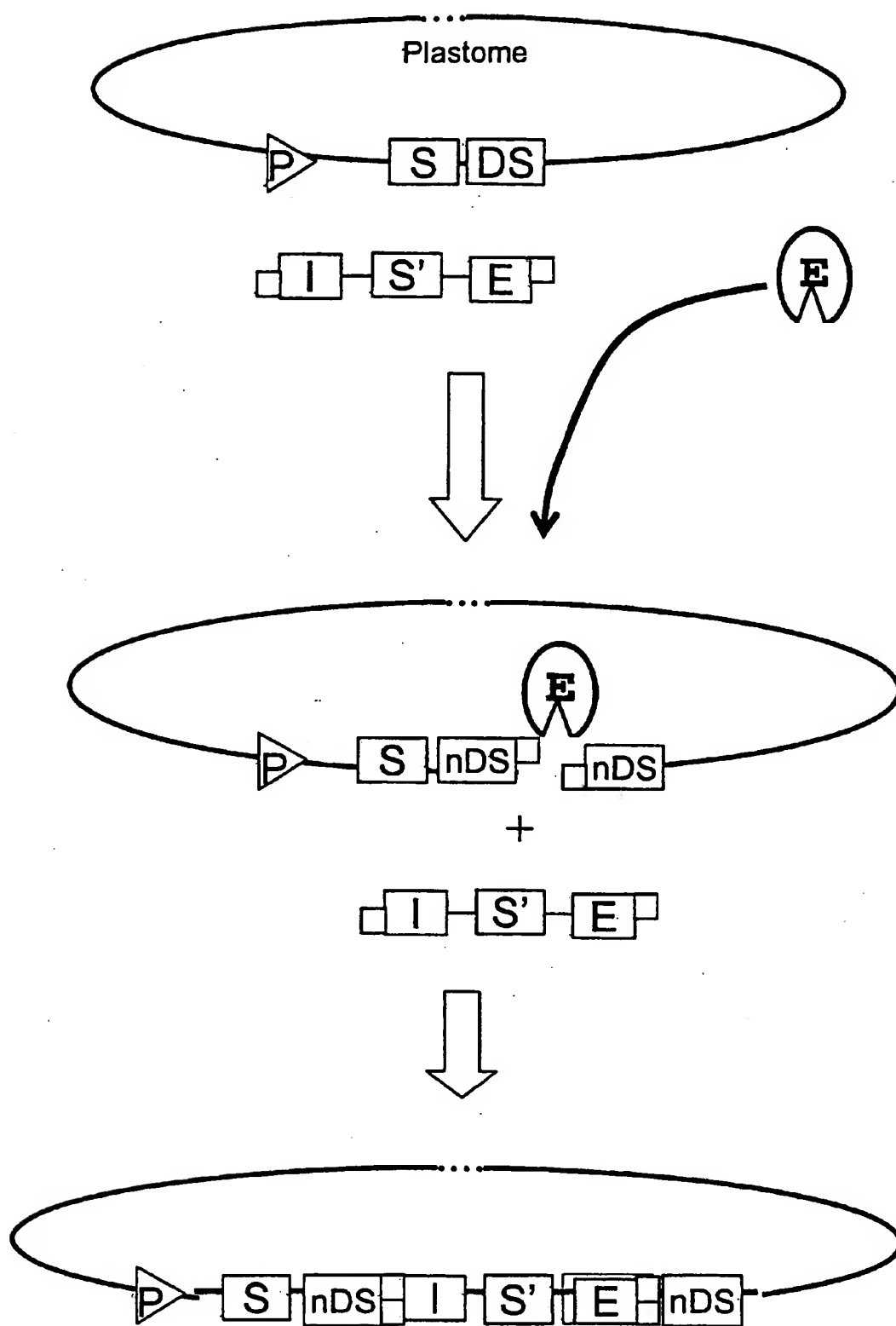


Fig. 6E

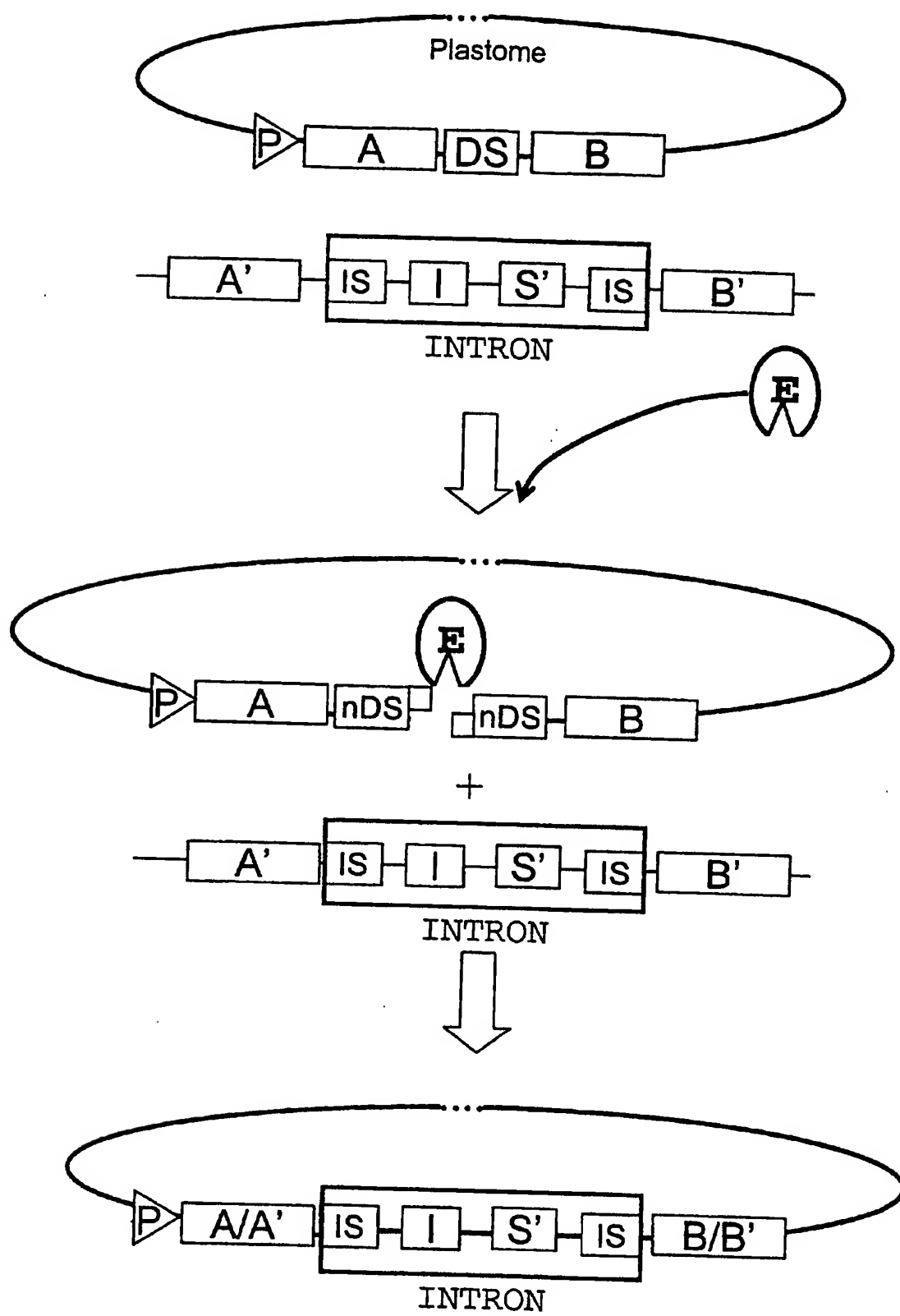


Fig. 7A



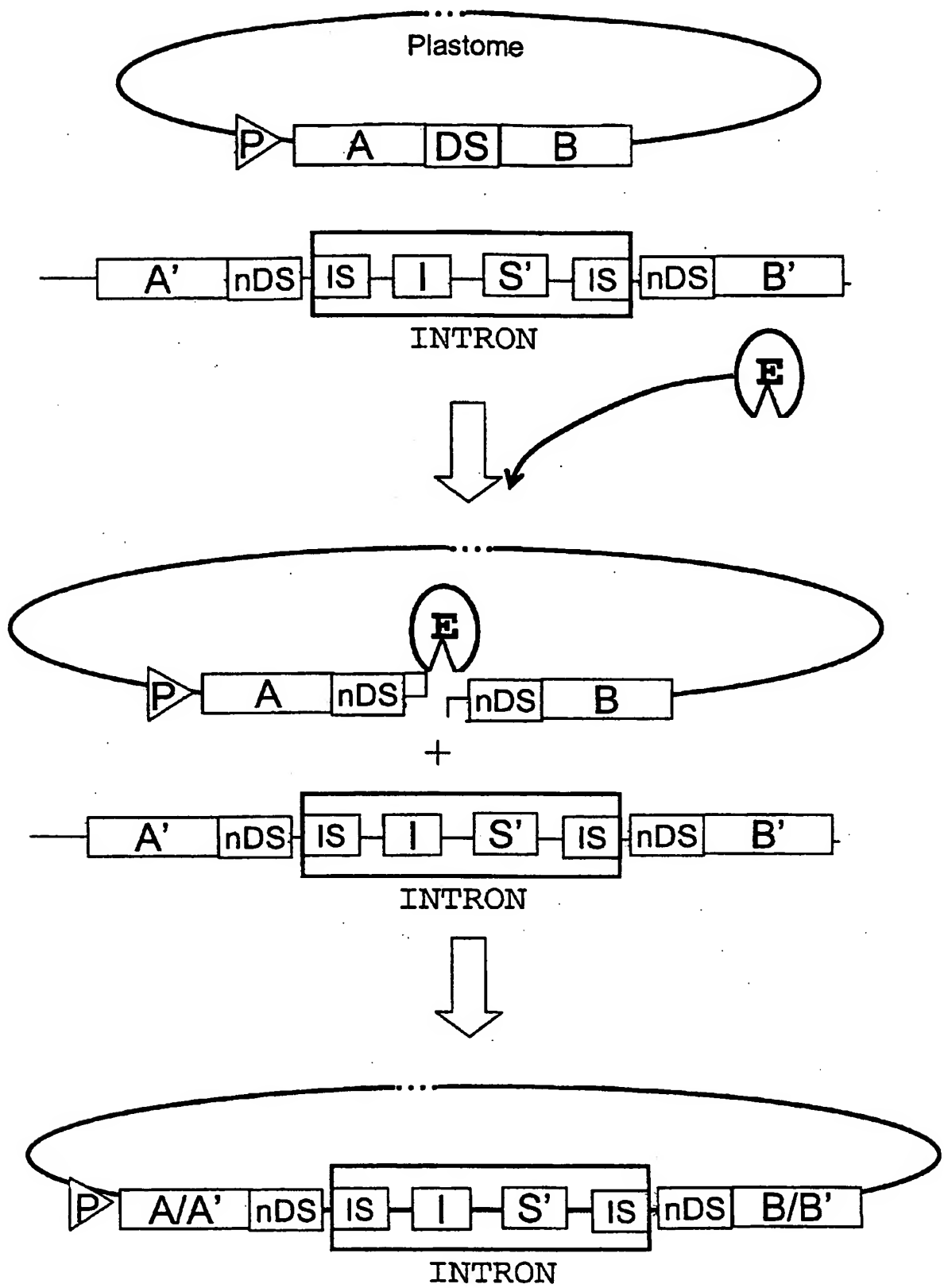


Fig. 7B

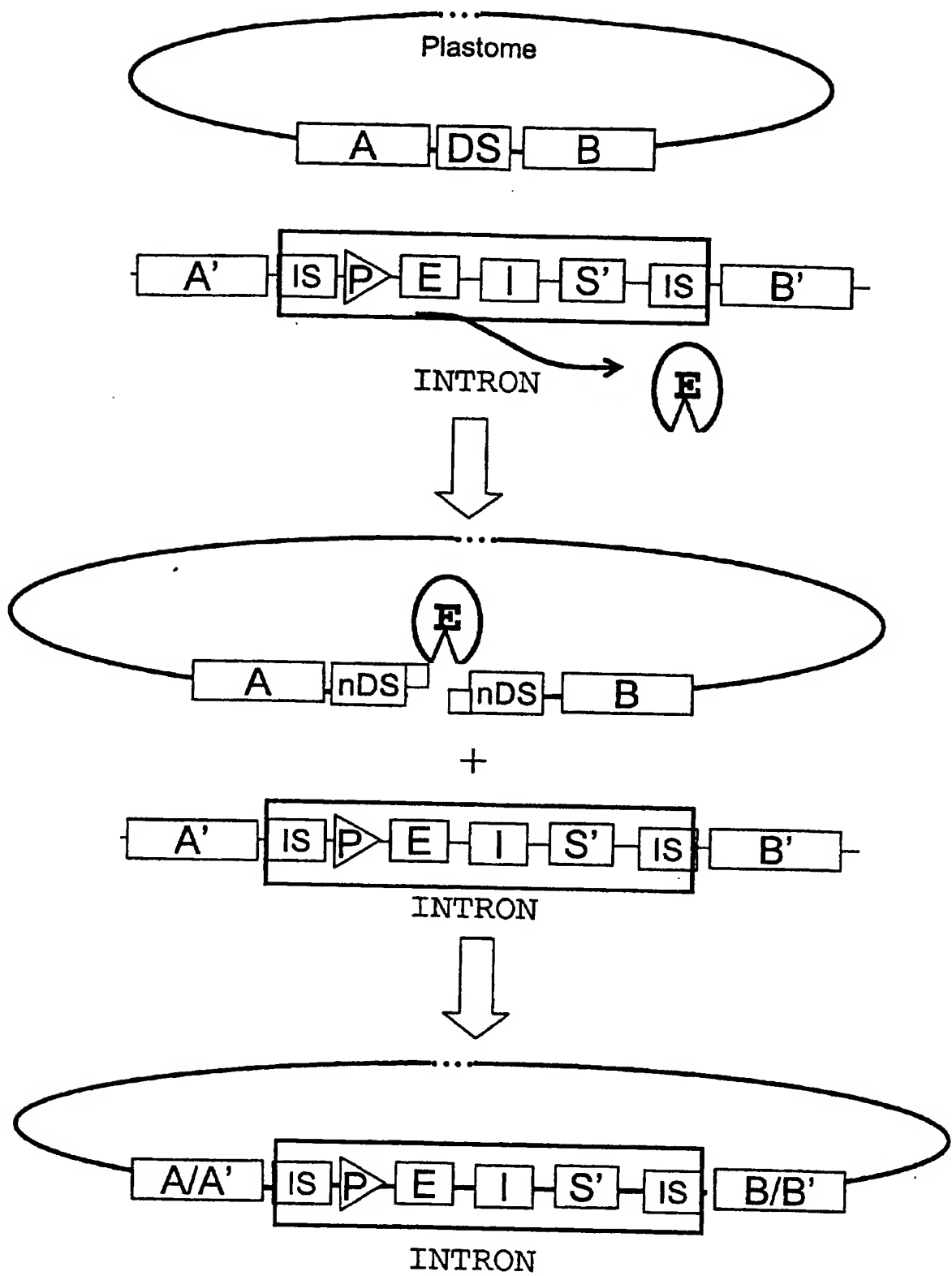


Fig. 7C

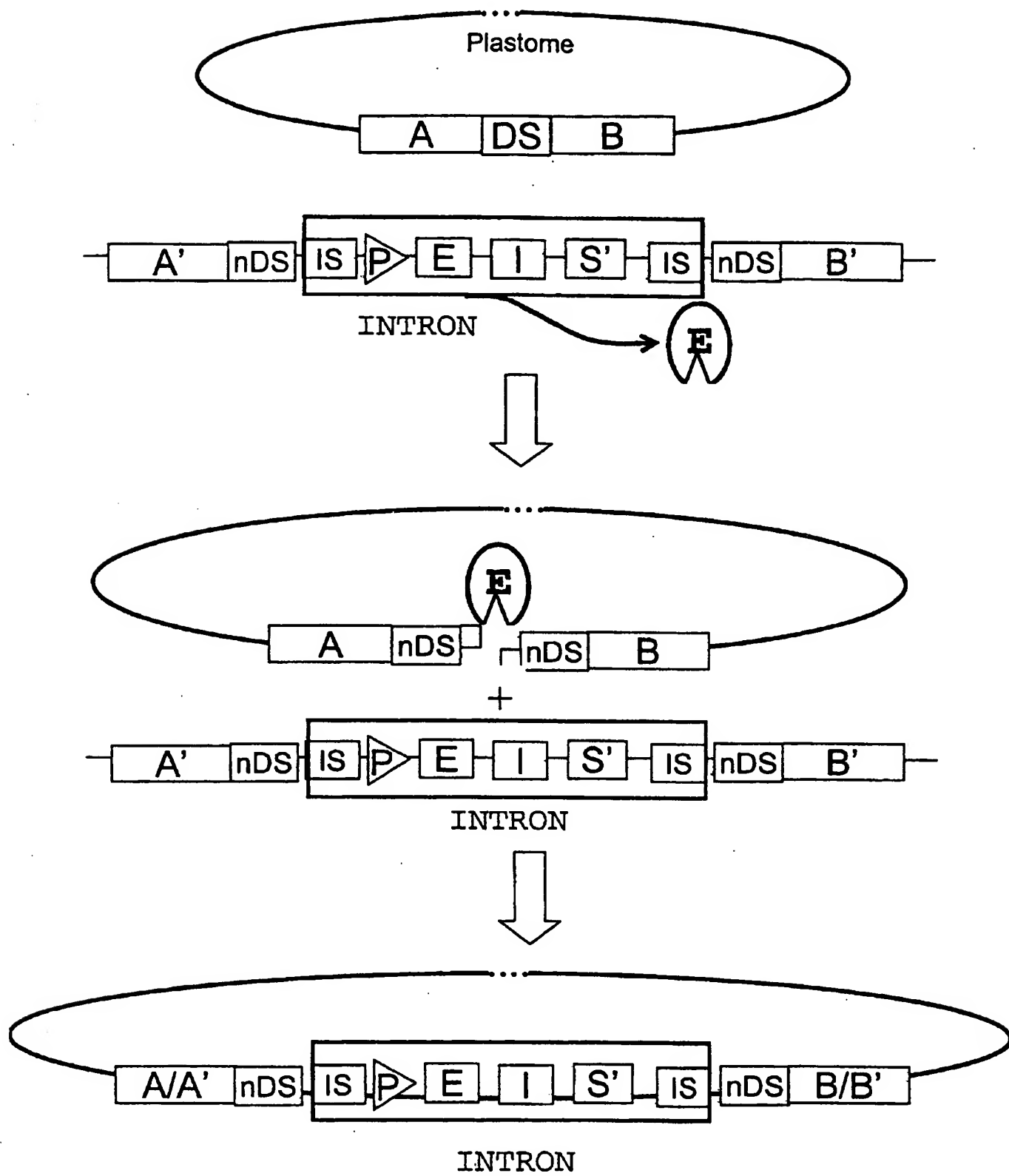


Fig. 7D

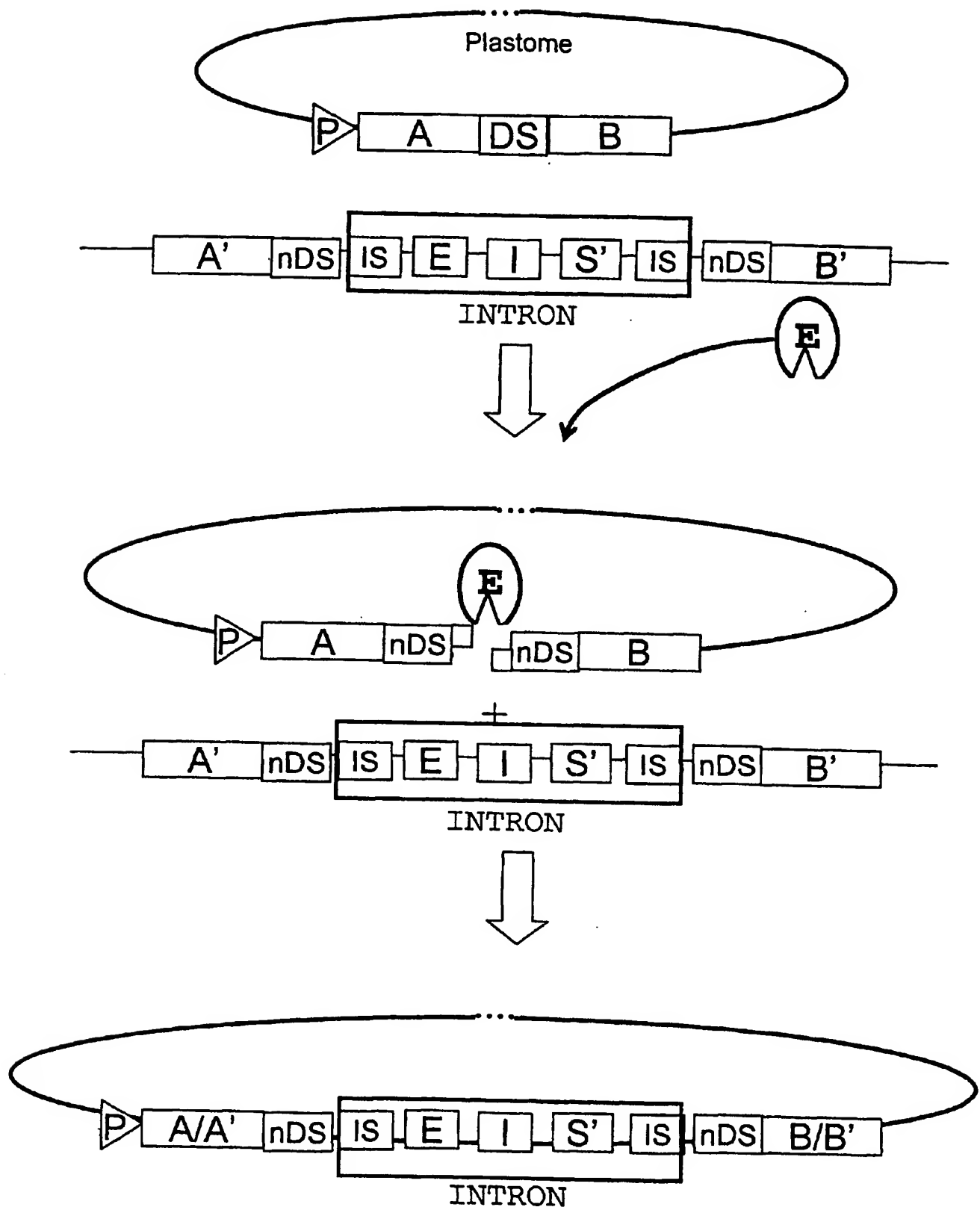


Fig. 7E

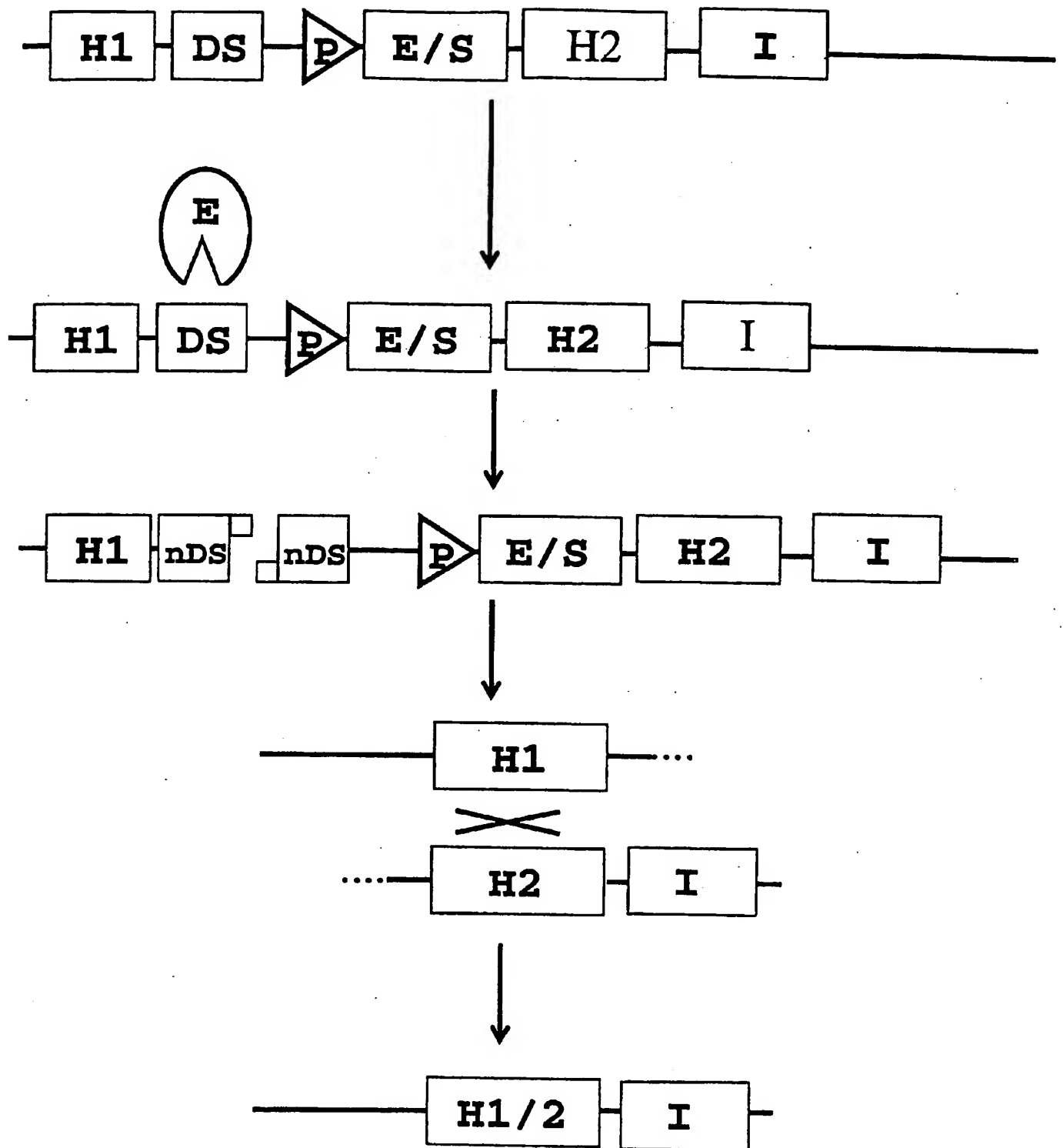
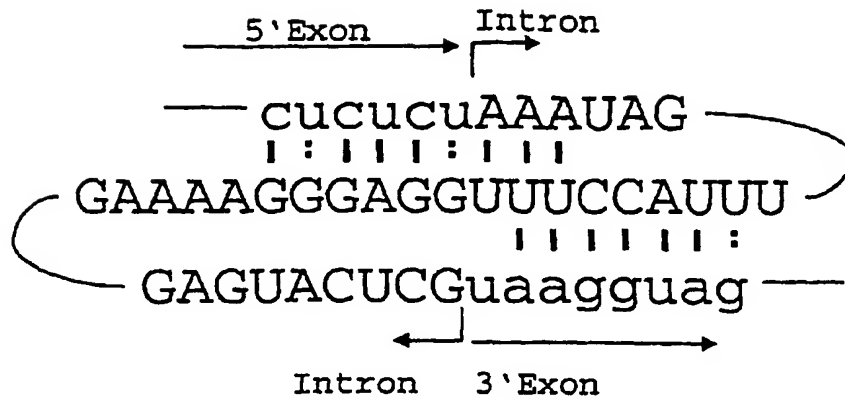


Fig. 8

A



B

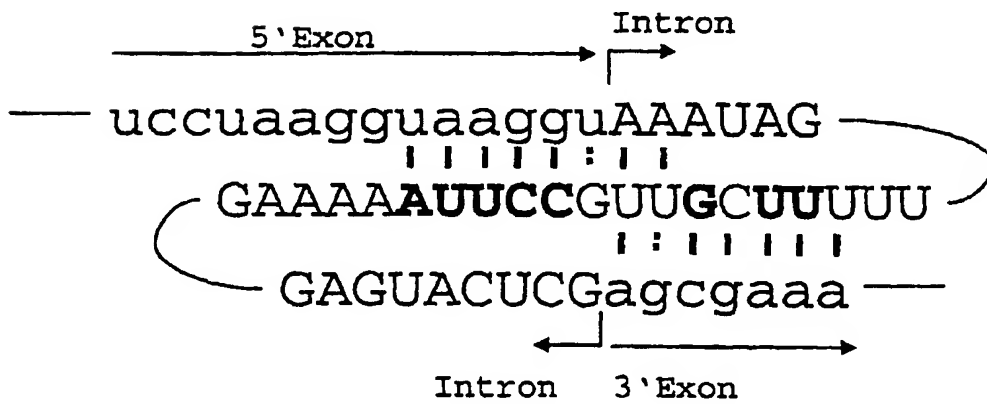


Fig. 10